

2011/21



Hierarchical shrinkage priors for dynamic regressions  
with many predictors

Dimitris Korobilis



**CORE**

DISCUSSION PAPER

Center for Operations Research  
and Econometrics

Voie du Roman Pays, 34  
B-1348 Louvain-la-Neuve  
Belgium

<http://www.uclouvain.be/core>

CORE DISCUSSION PAPER  
2011/21

**Hierarchical shrinkage priors for dynamic regressions  
with many predictors**

Dimitris KOROBILIS<sup>1</sup>

May 2011

**Abstract**

This paper builds on a simple unified representation of shrinkage Bayes estimators based on hierarchical Normal-Gamma priors. Various popular penalized least squares estimators for shrinkage and selection in regression models can be recovered using this single hierarchical Bayes formulation. Using 129 U.S. macroeconomic quarterly variables for the period 1959 – 2010 I exhaustively evaluate the forecasting properties of Bayesian shrinkage in regressions with many predictors. Results show that for particular data series hierarchical shrinkage dominates factor model forecasts, and hence is a valuable addition to existing methods for handling large dimensional data.

**Keywords:** forecasting, shrinkage, factor model, variable selection, Bayesian LASSO.

**JEL Classification:** C11, C22, C52, C53, C63, E37

---

<sup>1</sup> Université catholique de Louvain, CORE, B-1348 Louvain-la-Neuve, Belgium.  
E-mail: dimitrios.korompilis@uclouvain.be.

This paper presents research results of the Belgian Program on Interuniversity Poles of Attraction initiated by the Belgian State, Prime Minister's Office, Science Policy Programming. The scientific responsibility is assumed by the author.

## I Introduction

Responding to the vastly increasing need of applied economists in business and government for accurate economic forecasts using as much information as possible, academic econometricians have recently devoted significant effort to develop and test various methods for handling large macroeconomic and financial datasets. For many years, the dynamic factor model of Geweke (1977) has been used to successfully address the problem of summarizing datasets with hundreds of variables. In that respect, Stock and Watson (2002a,b), among many others, show that estimating dynamic factors, or just extracting principal components, can improve forecasts over simple ARMA models and also more complicated nonlinear time series models. In the last few years, many new statistical methods have emerged that do not explicitly summarize all the information in large datasets. Rather, they shrink their dimension by reducing or completely removing the impact of irrelevant predictors. These methods include statistical algorithms adopted in econometric forecasting, such as bagging (Inoue and Kilian, 2008), least absolute shrinkage and selection operator (LASSO) (De Mol, Giannone, and Reichlin, 2008), boosting (Bai and Ng, 2007), Bayesian model averaging (Koop and Potter, 2004) and dynamic model averaging (Koop and Korobilis, 2009).

More recently, Stock and Watson (2011) provide a flexible shrinkage representation of dynamic regression models with many orthogonal predictors. Their results are encouraging because they show that a global representation of many shrinkage estimators is possible, including pretest methods, Bayesian model averaging, empirical Bayes, and bagging. Their contribution is twofold since the theoretical properties of various shrinkage methods presented in previous literature depend on the specific modelling assumptions made, and empirical differences in the performance of these shrinkage methods rely on the data and implementation details of each study. Similarly, De Mol, Giannone, and Reichlin (2008) compare in one integrated setting the shrinkage and model selection properties of Bayesian LASSO and ridge regression estimators as opposed to principal component shrinkage. Both sets of authors identify that there is large potential in forecasting performance by shrinking the coefficients of a large number of predictors.

From a Bayesian point of view, the idea of a unified approach to shrinkage is not new. Long ago Bayes and Empirical Bayes priors which lead to shrinkage posterior estimators have been used successfully, with probably the most notable example in economics being the Minnesota prior for vector autoregressions of Litterman (1979) and the g-prior of Zellner. Empirical Bayes estimators in particular depend on a few hyperparameters which control the amount of shrinkage of each regression coefficient based on some information in the data sample. Additionally, early research identified the connection of Empirical Bayes methods with admissible estimators which dominate unrestricted simple least squares, like the Bayesian variant of the James-Stein estimator (see the results of Efron and Morris, 1975). Nowadays, modern Markov Chain Monte Carlo (MCMC) simulation methods can be used to consider a larger set of regularization ill-posed regression problems. In particular, the stochastic form of MCMC methods can be used to provide *adaptive* shrinkage and recover many estimators which may dominate least squares in a mean-square error sense.

In this paper I provide a quite general representation of Bayesian variants of penalized regression estimators. I show that by using hierarchical Normal-Gamma priors many popular estimators can be recovered such as the LASSO (Tibshirani, 1996), fused LASSO (Tibshirani et

al., 2005) and Elastic Net (Zou and Hastie, 2005). These priors are straightforward extensions of the typical Bayesian ridge regression priors used in regression models (Koop, 2003), and simple posterior expressions are available. In fact, following Kyung et al. (2010) the Bayes *posterior mode* of regression coefficients  $\beta$  in the typical regression problem  $y = \beta x + \varepsilon$  under a hierarchical Normal-Gamma prior admits a single general form which corresponds to the solution to a “generic” penalized regression problem of the form

$$\tilde{\beta}^{BAYES} = \arg \min_{\beta} \left\{ \|y - \beta x\|_{\ell_2} + \lambda_1 \|h_1(\beta)\|_{\ell_1} + \lambda_2 \|h_2(\beta)\|_{\ell_q} \right\},$$

where  $\|\cdot\|_{\ell_p}$  denotes the  $\ell_p$ -norm. This representation restricts the least squares estimator by adding two penalty terms  $\lambda_1 \|h_1(\beta)\|_{\ell_1}$  and  $\lambda_2 \|h_2(\beta)\|_{\ell_q}$ . Various specifications of the Normal-Gamma prior correspond to specific choices of  $h_1(\cdot)$ ,  $h_2(\cdot)$  and  $q$ , as well as the use of one or two regularization parameters  $\lambda_1$ ,  $\lambda_2$ . The benefit of a Bayesian approach using MCMC is that it is trivial to treat uncertainty about the regularization parameters  $\lambda_1$ ,  $\lambda_2$ , as well as relax the assumption of using the same amount of shrinkage for each regression coefficient (“adaptive shrinkage”) to obtain the oracle property (Zou, 2006).

The main goal of this paper is to empirically examine the shrinkage performance of the Normal-Gamma Bayes estimators using a data set with 129 quarterly macroeconomic and financial time series. For that reason I focus on five special cases of shrinkage estimators from the Normal-Gamma family and I set near improper (uninformative) priors on the regularization parameters  $\lambda_1$ ,  $\lambda_2$  as a default automatic choice. As Park and Casella (2008) note, scale invariance is not a compelling criterion for these parameters because they are unitless. However the purpose of this paper is to examine from a “practitioner’s point of view” if such automatic uninformative prior choices make sense for macroeconomic forecasting. In that case the frequentist econometrician can view hierarchical Bayes shrinkage as a pragmatic device and a useful tool for statistical inference (see for example the popularity of Bayesian model averaging in macroeconomic forecasting; Koop and Potter, 2004). The paper concludes with an application of shrinkage on forecasting GDP using many predictors focusing only on the Bayesian LASSO estimator. In this case I also perform a sensitivity analysis and compare the uninformative priors on the regularization parameters with some informative values (selected “subjectively”), as well as a semi-automatic method to estimate the regularization parameters based on marginal maximum likelihood.

The paper is organized as follows. Section 2 describes the econometric methodology: the general dynamic regression problem with many predictors; a unified shrinkage representation of Bayes estimators; their tuning; and how they compare with traditional shrinkage. Section 3 reports the results from the out-of-sample exercise for five special shrinkage estimators applied on 129 series. This section concludes with a sensitivity analysis. Section 4 concludes and provides an assessment of the empirical value of hierarchical shrinkage priors.

## II Bayes shrinkage formulations for dynamic regressions

In this paper I consider univariate forecasting models of the form

$$y_{t+h} = \alpha z_t + \beta x_t + \varepsilon_{t+h}, \quad (1)$$

where  $\varepsilon_{t+h}$  is the  $h$ -quarters ahead forecasts error distributed  $\varepsilon_t \stackrel{iid}{\sim} N(0, \sigma^2)$ , for  $t = 1, \dots, T$ . In this type of regressions  $y_{t+h}$  is the  $h$ -quarters ahead value of the variable of interest,  $z_t$  is the  $q \times 1$  vector of unrestricted predictors always included in the forecasting model, like the intercept, dummies and lags of the dependent variables, and  $x_t$  is the  $p \times 1$  vector of *many* (say  $p \rightarrow \infty$  or  $p$  grows at a faster rate than  $T$ ) *standardized* exogenous predictors whose dimension we would like to shrink.

The unrestricted coefficients  $\alpha$  and the variance  $\sigma^2$  can be integrated out using the uninformative priors  $\pi(\alpha) \propto 1$  and  $\pi(\sigma^2) \propto 1/\sigma^2$  respectively. This allows closer focus on the regression coefficient vector  $\beta$  which has individual elements  $\beta_j, j = 1, \dots, p$ .

### II.1 Classical shrinkage

A noninformative prior, like the one assigned to the coefficients  $a$ , leads to a Bayes estimator centered at the unrestricted LS quantities. This choice would obviously pose a problem for estimating the “large” number of coefficients  $\beta$ , especially when  $p > T$ . Traditionally, Normal priors of the form

$$\pi(\beta) \sim N(0, V), \quad (2)$$

have been used, because they are conjugate to the likelihood and allow easy calculations of the Bayes posterior. The  $p \times p$  matrix  $V$  is the prior covariance matrix of the regression coefficients which we want to elicit for this “large  $p$ ” problem. For instance, a common choice is the case  $V = \tau^2 I_p$  which leads to the classical *ridge regression* shrinkage. Ignoring for now the effect of the regressors  $z_t$ , this ridge regression prior implies the penalized least squares representation

$$\beta = \left( X'X + \frac{1}{\tau^2} I_p \right)^{-1} X'Y$$

where  $X = (x'_1, \dots, x'_T)'$  and  $Y = (y'_{1+h}, \dots, y'_{T+h})'$ . The dependence of all parameters  $\beta_j, j = 1, \dots, p$ , on the unknown parameter  $\tau^2$  can reduce the risk over the traditional LS estimator. For  $\tau \rightarrow \infty$  we can see that  $\beta = (X'X)^{-1} X'Y = \beta^{LS}$ .

Following a different path, Judge and Bock (1978) suggested an Empirical Bayes (i.e. data-based) estimator of  $V$ , of the form  $V = \tau^2 (X'X)^{-1}$  where  $\tau^2 = \frac{\hat{\sigma}^2}{\hat{\xi}}$ , and

$$\begin{aligned} \hat{\sigma}^2 &= (Y - X\beta^{LS})' (Y - X\beta^{LS}) / T \\ \hat{\xi}^2 &= \frac{\beta^{LS'} \beta^{LS}}{\text{tr}(X'X)^{-1}} - \hat{\sigma}^2 \end{aligned}$$

This empirical Bayes rule is Stein-like, shrinking  $\beta^{LS}$  towards 0, since the posterior mean writes

$$\tilde{\beta} = \left(1 - \frac{\hat{\sigma}^2}{\hat{\sigma}^2 + \hat{\xi}^2}\right) \beta^{LS}.$$

Nowadays, priors of the form  $V = \sigma^2 \tau^2 (X'X)^{-1}$ , which are called  $g$ -priors or Zellner's prior (Zellner, 1986), tend to be very popular in economics; see for instance Koop and Potter (2004) and references therein. Over the course of the years there have been many connections between values of  $\tau^2$  and information criteria; see for example Fernandez, Ley and Steel (2001) for a review. Nevertheless, Zellner originally proposed this prior to provide a shrinkage representation since the posterior mean writes

$$\beta = \frac{\tau^2}{1 + \tau^2} \beta^{LS}$$

This formulation implies a shrinkage factor  $\delta = \tau^2 / (1 + \tau^2)$  which regulates the proportion (0%-100%) of shrinkage over the unrestricted OLS estimator. Note that priors which are data-based have desirable shrinkage properties, compared to the weak shrinkage of a ridge-regression prior.

We can immediately observe that these two typical examples of Bayesian shrinkage have undesirable properties for very demanding problems with many predictors. The ridge regression prior is based on a global shrinkage parameter  $\tau^2$  for all  $p$  regressors. In sparse regression problems, i.e. when  $p$  is very large and we expect that only a tiny proportion of regressors are relevant for prediction, weighting a-priori all regression coefficients by the same factor  $\tau^2$  is guaranteed not to work well. Empirical Bayes priors partly solve this problem since  $\tau^2$  is scaled by the Information Matrix, giving a varying degree of prior weight to each regression coefficient based on the information in the likelihood. Nevertheless, the Information Matrix cannot be estimated precisely (for large  $p$ ), or cannot be estimated at all (for  $p > T$ )<sup>1</sup>. Thus, the next subsection develops on shrinkage representations which are automatic (i.e. they allow minimal input by the researcher about the expected shrinkage factor) and can be applied in sparse regressions within the “large  $p$ , small  $T$ ” paradigm.

## II.2 Full Bayes (hierarchical) priors for adaptive shrinkage

Modern computational methods allow to estimate the parameter(s) in the prior covariance matrix  $V$  in a formal way, by placing hyper-prior distributions on these parameters. Moreover, adding an extra layer of hierarchy on the prior covariance matrix (and hence treating this matrix as a parameter to be estimated from the likelihood) allows to implement many popular formulations of adaptive shrinkage. For that reason, the prior covariance matrix on the coefficients  $\beta$  is formulated as  $V = \text{diag} \{ \tau_1^2, \dots, \tau_p^2 \}$ . This formulation allows the individual elements  $\tau_j^2$ ,  $j = 1, \dots, p$ , to be *independently* updated towards 0 which eventually results in shrinkage of the coefficient  $\beta_j$  to a point mass at zero. All hierarchical priors presented below are special

---

<sup>1</sup>It is only recently that Maruyama and George (2010) derived a particular decomposition of Zellner's  $g$ -prior that can be used when more predictors than observations are present.

cases of a Normal-Gamma prior, i.e. a Normal prior for  $\beta$ , and a Gamma prior for  $\tau_j^2$  of the form

$$\begin{aligned}\pi(\beta|\tau^2) &\sim N_p(0, V) \\ \pi(\tau_j^2) &\sim \text{Gamma}(a, b)\end{aligned}$$

This formulation is very flexible and nests many cases used previously in the shrinkage literature. Given the properties of the Gamma distribution I will give special attention to the cases

1.  $\tau_j^2 \sim \text{Gamma}(a = 0^+, b = 0^+)$  which is equivalent to  $\log(\tau_j^2) \sim \text{Uniform}^*[0, +\infty)$
2.  $1/\tau_j^2 \sim \text{Gamma}(a = \rho, b = \xi)$  which is equivalent to  $\tau_j^2 \sim i\text{Gamma}(\rho, \xi)$ , and
3.  $\tau_j^2 \sim \text{Gamma}(a = 1, b = \frac{2}{\lambda^2})$  which is equivalent to  $\tau_j^2 \sim \text{Exponential}(\frac{\lambda^2}{2})$

where  $\text{Uniform}^*$  denotes the unnormalized (and hence improper) Uniform distribution, and  $i\text{Gamma}$  is the inverse Gamma distribution.

These hierarchical priors basically transform the typical independent Normal prior in (2) into a scale mixture of Normals prior. These examples can basically be generalized to even more cases. However, for the specific choices made above, various known distributions can be recovered. For instance, case 2 (the Normal-inverse Gamma density) is a mixture representation for the t-density (Geweke, 1993), while case 3 (the Normal-Exponential density, also called “double-exponential” density) is a mixture representation of the Laplace density. Figure 1 shows intuitively why a mixture prior (right panel) is to be preferred over the traditional ridge regression prior (left panel). While the ridge regression prior is informative on the support of the parameters (it is bell-shaped, as opposed to being completely flat like a uniform prior), it is locally uninformative in a neighborhood zero (the point of shrinkage). In contrast, the Laplace prior (which is constructed in such a way that it has similar support to the ridge regression prior) provides faster rates of shrinkage in the neighborhood of zero.

**Adaptive shrinkage Jeffreys’ prior** Hobert and Casella (1996) studied first the shrinkage properties of the Jeffreys’ prior on the covariance matrix of the regression coefficients. One can think of Jeffreys’ prior as the simplest, default choice because it is not dependent upon further hyperparameters.

Let  $V = \text{diag}\{\tau_1^2, \dots, \tau_p^2\}$ , then the scale-invariant, improper Jeffrey’s (hyper-)prior on each  $\tau_j^2$  takes the form

$$\pi(\tau_j^2) \sim 1/\tau_j^2, \text{ for } j = 1, \dots, p \quad (3)$$

**Adaptive shrinkage t-priors** For a covariance matrix  $V = \text{diag}\{\tau_1^2, \dots, \tau_p^2\}$  we can consider a specific form of a Gamma prior on  $\tau_j^2$ ,  $j = 1, \dots, p$ , i.e. the inverse Gamma prior. Following Geweke (1993) we can show that this Normal-inverse Gamma mixture prior is equivalent to a Student- $t$  prior on  $\beta$ . The  $t$ -density has heavy tails and is more leptokurtic around the origin, but in general is much “smoother” than the Laplace density plotted in panel B of Figure 1. The

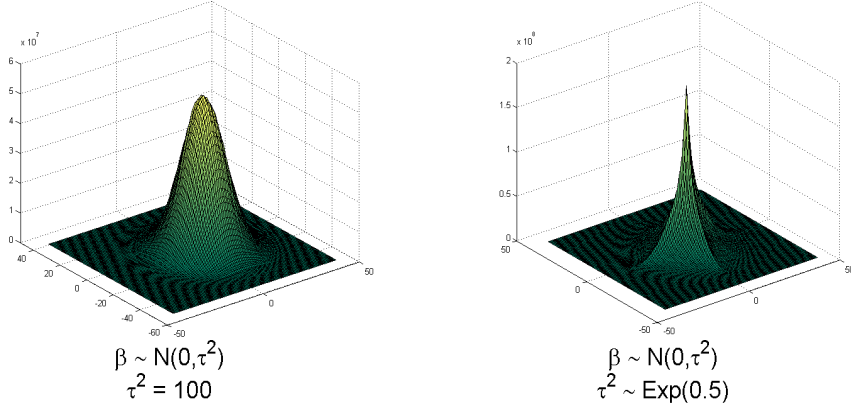


Figure 1: The left panel of the figure plots the ridge regression prior which is informative on the support of the parameters  $\beta$ . For large  $\tau$ , this prior is locally uninformative (flat) around zero and has no shrinkage properties. The right panel shows how the shrinkage towards zero is achieved using the Laplace prior.

priors on  $\tau_j^2$  are of the form

$$\pi(\tau_j^2) \sim iGamma(\rho, \xi), \text{ for } j = 1, \dots, p \quad (4)$$

where  $\rho$  is the *shape* parameter and  $\xi$  the *scale* parameter of the inverse Gamma density; see also Armagan and Zaretzki (2010). Once the  $\tau_j^2$ 's are integrated out from the joint posterior, this prior is analogous to the regularized least squares problem which solves (ignoring once again the regressors  $z_t$  for simplicity)

$$\arg \min_{\beta} \frac{1}{2\sigma^2} \sum_{t=1}^T (y_{t+h} - x_t \beta)^2 + \left( \rho + \frac{1}{2} \right) \sum_{j=1}^p \log(2\xi + \beta_j)$$

Finally, notice that this formula also applies for the Jeffrey's prior case (for  $\rho, \xi \rightarrow 0$ ).

**Hierarchical LASSO** Tibshirani (1996) proposed the Lasso algorithm which can be viewed as a  $L_1$ -penalized least squares estimate which solves

$$\arg \min_{\beta} \frac{1}{2\sigma^2} \sum_{t=1}^T (y_{t+h} - x_t \beta)^2 + \lambda \sum_{j=1}^p |\beta_j|$$

Tibshirani (1996) also noted that this form of penalty is equivalent to the posterior mode of the Bayes estimate under the Laplace prior

$$\pi(\beta|\sigma^2) \sim \prod_{j=1}^p \frac{\lambda}{2\sqrt{\sigma^2}} e^{-\frac{\lambda}{\sqrt{\sigma^2}}|\beta_j|}$$



One can take advantage of the fact that the Laplace density can be written as a scaled mixture of Normals (see Park and Casella, 2008). Notice that the formulation above implies that for the Bayesian LASSO prior (as well as the Fused LASSO and the Elastic Net) we need to condition on the error variance  $\sigma^2$ . Park and Casella (2008) underline that this conditioning ensures that the posterior of the regression coefficients  $\beta$  is unimodal, otherwise expensive simulation methods would be needed to handle multimodal posteriors (for instance, simulated tempering). Subsequently, assume for this case a diagonal prior covariance matrix of the form  $V = \sigma^2 \times \text{diag} \{ \tau_1^2, \dots, \tau_p^2 \}$ . The hierarchical version of the LASSO uses a normal prior for  $\beta$  of the form in eq. (2) augmented with the hyperprior

$$\pi(\tau_j^2) \sim \text{Exponential} \left( \frac{\lambda^2}{2} \right), \text{ for } j = 1, \dots, p \quad (5)$$

where  $\lambda$  is a hyperparameter, which is the *rate* parameter of the Exponential distribution.

**Hierarchical Fused LASSO** The Fused LASSO was proposed by Tibshirani et al. (2005) as a means to account for a possible meaningful ordering of variables<sup>2</sup>. Thus, this estimator penalizes the  $L_1$ -norm of both the coefficients and their differences

$$\arg \min_{\beta} \frac{1}{2\sigma^2} \sum_{t=1}^T (y_{t+h} - x_t \beta)^2 + \lambda_1 \sum_{j=1}^p |\beta_j| + \lambda_2 \sum_{j=1}^{p-1} |\beta_{j+1} - \beta_j|$$

The representation of the Bayesian prior for  $\beta$  in the penalized regression using the Fused LASSO is

$$\pi(\beta | \sigma^2) \sim e^{-\frac{\lambda_1}{\sigma} \sum_{j=1}^p |\beta_j| - \frac{\lambda_2}{\sigma} \sum_{j=1}^{p-1} |\beta_{j+1} - \beta_j|}$$

Kyung et al. (2010) show that the hierarchical representation of this prior is

$$\pi(\tau_j^2) \sim \text{Exponential} \left( \frac{\lambda_1^2}{2} \right), \text{ for } j = 1, \dots, p \quad (6a)$$

$$\pi(\omega_j^2) \sim \text{Exponential} \left( \frac{\lambda_2^2}{2} \right), \text{ for } j = 1, \dots, p-1 \quad (6b)$$

where the correlation between  $\beta_{j+1}$  and  $\beta_j$  enters through the prior covariance matrix  $V$ . In this case  $V$  is a tridiagonal matrix with main diagonal  $\{\tau_i^2 + \omega_{i-1}^2 + \omega_i^2\}$  for  $i = 1, \dots, p$  and off-diagonal elements  $\{-\omega_i^2\}$ , and for simplicity we can set  $\omega_0 = \omega_p = 0$ .

**Hierarchical Elastic Net** Zou and Hastie (2005) proposed the Elastic Net as a more stabilized version of the LASSO that also allows grouping effects and is particularly useful when  $p > T$ .

---

<sup>2</sup>The data set in this paper implies such an ordering. Many disaggregated and component series of the same aggregated series appear in order. Additionally, all variables in this dataset are ordered according to statistical releases.

The Elastic Net estimator is the solution to the minimization problem

$$\arg \min_{\beta} \frac{1}{2\sigma^2} \sum_{t=1}^T (y_{t+h} - x_t\beta)^2 + \lambda_1 \sum_{j=1}^p |\beta_j| + \lambda_2 \sum_{j=1}^p \beta_j^2$$

A Bayesian prior for  $\beta$  in the penalized regression using this estimator is

$$\pi(\beta|\sigma^2) \sim e^{-\frac{\lambda_1}{\sqrt{\sigma^2}} \sum_{j=1}^p |\beta_j| - \frac{\lambda_2}{2\sigma^2} \sum_{j=1}^p \beta_j^2}$$

Kyung et al. (2010) show that a hierarchical representation of this density exists, and it is of double-exponential form, as in the simple LASSO. This means that the hyperprior on  $\tau_j^2$  is

$$\pi(\tau_j^2|\lambda_1^2) \sim \text{Exponential}\left(\frac{\lambda_1^2}{2}\right), \text{ for } j = 1, \dots, p \quad (7a)$$

where in this case the difference with the standard LASSO prior is that the covariance matrix is of the form  $V = \sigma^2 \times \text{diag}\left\{(\tau_1^{-2} + \lambda_2)^{-1}, \dots, (\tau_p^{-2} + \lambda_2)^{-1}\right\}$ .

As opposed to maximizing the likelihood using no prior information, estimation for the Bayesian means that the likelihood function has to be averaged using each of the five priors presented above. This weighted average is the posterior distribution of the regression coefficients  $\beta$ , and the mode of the posterior is identical to frequentist shrinkage estimators. In Appendix B I give all the necessary details on how to get samples from the posterior distribution of all regression coefficients  $\theta = \{\alpha, \beta, \sigma^2\}$  by sampling from their conditional posteriors using Markov Chain Monte Carlo (MCMC) methods. These are fairly easy to implement and computationally efficient.

It should be noted here that other extensions of these priors are possible, although many of these extensions can become cumbersome computationally. For instance one can use the fact that a *Gamma* ( $a = v/2, b = 2$ ) distribution is equivalent to a  $\chi^2 \sim (v)$  distribution, with the cost that a Normal-Chi-square mixture is not a representation of any known distribution with known desirable shrinkage properties. Similarly, Park and Casella (2008) discuss some alternative priors based specifically on the LASSO, such as the extension proposed by Rosset and Zhu (2004). These authors propose to robustify the LASSO by considering a quadratic Huber-type loss function  $H$  which has the property that the coefficients  $\beta$  are shrunk quadratically around zero, while outside the neighborhood of zero this function becomes piecewise linear. This ‘‘Huberized LASSO’’ takes the form

$$\min_{\beta} \sum_{t=1}^T H(y_{t+h} - x_t\beta) + \lambda \sum_{j=1}^p |\beta_j|$$

but Park and Casella (2008) note that in this case it is not straightforward to marginalize over  $\alpha z_t$  (which was purposely ignored in our discussion so far, since as Appendix B shows, it is easy to marginalize over  $\alpha z_t$  when assuming the five priors presented above). Lastly, Hobert and Geyer (1998) proved geometric ergodicity of the two-stage Gibbs sampler from hierarchical

models of a general Normal-Gamma form, a result which can be generalized to the LASSO, Fused LASSO and Elastic Net priors (see Kyung et al., 2010).

### II.2.1 Tuning the hyperparameters

Hierarchical priors provide the advantage of allowing the data to determine the prior hyperparameter of interest (covariance of the Normal prior in our case). However from the formulations above we can observe that introducing a second layer of hierarchy (the Gamma-type densities) means that at least one new hyperparameter is introduced; it is only for the Normal-Uniform prior that this is obviously not the case. For the Normal-inverse Gamma prior (Student- $t$ ) we need to select values for the hyperparameters  $(\rho, \xi)$  of the inverse Gamma density. Although one can easily set a prior on the scale parameter  $\xi^3$ , typical uninformative values for the inverse Gamma distribution in Bayesian analysis are usually  $\rho = \xi = 0.01$  or  $\rho = \xi = 0.001$  (see Gelman, 2006). Since for these very low values of  $(\rho, \xi)$  the inverse Gamma becomes equivalent to a Jeffrey's prior for  $\tau_j^2$  (which is the first shrinkage prior examined), I will examine the more informative prior  $iGamma(\rho = 3, \xi = 0.001)$  which concentrates  $\tau_j^2$  around the neighborhood of zero (note that for  $\rho \leq 2$  the variance of the inverse Gamma does not exist).

In the Hierarchical LASSO prior case, in the absence of other information, we should find uninformative values for the Exponential prior distribution of  $\tau_j^2$ . Therefore we would want to make a specific choice of the rate parameter  $\lambda$  that would give a combination of a low prior mean value for  $\tau_j^2$  (ideally zero) and a quite large prior variance. Given that the mean and variance of an *Exponential*( $\lambda$ ) distribution are  $\lambda^{-1}$  and  $\lambda^{-2}$  respectively, this is not a straightforward combination to achieve (both mean and variance increase or decrease at the same time). In that respect, one can introduce an additional hierarchical layer for the parameters  $\lambda$ . A conjugate prior which would facilitate posterior computations when using the Exponential prior, is the Gamma prior on  $\lambda^2$  (not  $\lambda$ ) of the form

$$\pi(\lambda^2) \sim Gamma(r, \delta)$$

Similarly, an additional layer on the hyperparameters  $\lambda_1, \lambda_2$  of the Fused Lasso and Elastic Net priors is of the form

$$\begin{aligned} \pi(\lambda_1^2) &\sim Gamma(r_1, \delta_1) \\ \pi(\lambda_2^2) &\sim Gamma(r_2, \delta_2) \end{aligned}$$

and hence now it easy to verify that setting  $r = \delta = 0.01$  (similarly  $r_1 = \delta_1 = r_2 = \delta_2 = 0.01$ ) we have a near-Uniform (noninformative) prior on the hyperparameters  $\lambda, \lambda_1, \lambda_2$ .

---

<sup>3</sup>A conjugate prior on  $\xi$  is the *Gamma*( $\alpha_0, \beta_0$ ) density. Then the posterior of the inverse Gamma prior is again an inverse Gamma density with parameters  $(\alpha_0 + p\rho, \beta_0 + \sum_{i=1}^p \tau_i^{-2})$ .

### III Empirical Results

The data-set consists of 129 quarterly U.S. macroeconomic time series spanning the period 1959:Q1 to 2010:Q2 (the effective sample size, after converting to stationary and taking lags is 1960:Q1-2010:Q2). The series were downloaded from the St. Louis Fed FRED database (<http://research.stlouisfed.org/fred2/>) and a complete description is given in Table A.1 in the Appendix. The whole dataset is quite standard for this type of application, and includes among others data releases like personal income and outlays, GDP and components, assets and liabilities of commercial banks in the United States, productivity and costs measures, exchange rates and selected interest rates. All series are seasonally adjusted, where this is applicable, and transformed to be approximately stationary. All transformations are summarized in column “T” in Table A.1 and explained in detail in Appendix A. Bottom line is that when the series are used as predictors in  $x_t$ , standard stationarity transformations are applied, like first and second (log) differences. In contrast, when the series are used as the series to be predicted ( $y_{t+h}$ ), then  $h$ -quarter growth or differences transformations are used.

In the dataset there are series which are higher level aggregates (mainly sums) of individual disaggregated series. There are 14 series like that in the dataset which are excluded when extracting factors, as it is not sensible to extract a common factor between, say, two series and their sum. Column “F” in Table A.1 denotes with 1 only the 115 disaggregated variables which are used to extract factors. This restriction does not hold when using the shrinkage priors and all series are used as predictors.

#### III.1 Forecasting with many predictors

All forecasts are from the univariate regression (1) where iteratively I use one of the 129 variables as the dependent variable ( $y_{t+h}$ ) and the remaining 128 variables enter the regression as the matrix of standardized exogenous predictors ( $x_t$ ). Then the five Bayesian shrinkage priors are applied to estimate  $\hat{\beta}^j$ ,  $j = \text{Jeffreys, Student-t, Lasso, Fused Lasso, Elastic Net}$ , and forecasts are produced using the original, unstandardized matrix of predictors  $x_t^*$ . In order to forecast with the dynamic factor model (DFM),  $x_t$  is replaced with the first five principal components of the 115 disaggregated series in  $x_t$  and  $\hat{\beta}^{DFM}$  is estimated with simple OLS. The variables which are always included in each of the six forecasting models ( $z_t$ ) are the intercept and two lags of the one-quarter growth rates or differences of the dependent variable (i.e. lags  $y_t$  and  $y_{t-1}$  using the same stationarity transformations as in the variables in  $x_t$ ).

The first estimation period is 1960:Q1 (after taking lags and transforming to stationarity) to 1984:Q4 and the sample 1985:Q1 to 2010:Q2 (last 102 observations) is kept for evaluation of  $h$ -step ahead forecasts,  $h = 1, 2, 4$ . In particular, using the initial sample (where  $y_{t+h}$  is observed from 1960:Q1 to 1984:Q4 and ( $z_t, x_t$ ) is observed from 1960:Q1 to 1984:Q4- $h$ ) estimation of the regression (1) provides parameter estimates  $\hat{\alpha}, \hat{\beta}, \hat{\sigma}^2$ , and then forecasts can be computed for  $y_{1984:Q4+h}$  by plugging-in the regression the realization of the predictors in 1984:Q4, i.e. the values ( $z_{1984:Q4}, x_{1984:Q4}$ ). Then one data point is added and the same procedure is followed until the sample is exhausted. Since the models with shrinkage priors are estimated using MCMC (see Appendix B), which provides draws from the whole posterior density of the parameters, predictive simulation is used to obtain the whole predictive density. For each of the 129 dependent

variables, the five prior distributions, the three forecast horizon, and the 102- $h$  out-of-sample observations, 7.000 post-burn in draws from the conditional posteriors of the regression parameters  $(\alpha, \beta, \sigma^2)$  are saved (see Appendix B for exact formulæ), and using each parameter draw 10 forecasts are generated leading in 70.000 draws from the predictive density of each of the 129 variables.

In a similar comparison of shrinkage estimators for regressions with many predictors, Stock and Watson (2011) use 4 lags in each of their 143 univariate regressions and report all their results relative to an AR(4) model. Del Mol, Giannone and Reichlin (2008) consider only an unrestricted intercept in their shrinkage regressions and report results relative to a random walk. In this paper, since the effects of an intercept and two lags are partialled out in each forecasting model, it is natural to consider forecast performance statistics relative to an AR(2) model. In this paper Mean Absolute Forecast Errors (MAFE) and Mean Squared Forecast Errors (MSFE) are considered. Unless stated otherwise, all results are based on the MAFE and MSFE statistics of model  $j$  relative to the MAFE and MSFE of the AR(2) model (i.e.  $MAFE_j = mafe_j / mafe_{AR(2)}$  and  $MSFE_j = msfe_j / msfe_{AR(2)}$ ). Consequently,  $MAFE_j > 1$  means that the AR(2) dominates in terms of absolute forecast error, while the opposite is true when  $MAFE_j < 1$ .

Tables 1 to 3 present the average absolute forecast errors for 1,2 and 4 quarters ahead. Since the relative MAFE results are averaged over many series, three decimals are used in this table because otherwise the differences are quite small (see also Stock and Watson, 2011). First, based on the median MAFE using the total number of series, the simple LASSO and the Elastic Net give the smallest forecast errors in all cases (note that for  $h = 2$  there is a difference but this is minimal). This might suggest that taking into account the correlation among the predictors, which is what the Elastic Net algorithm adds to the simple LASSO algorithm, is not that important with these data. However, the Elastic Net consistently has the smallest maximum MAFE, and consequently has the smallest variance across all 129 series. The same shrinkage algorithm clearly dominates in most of the 17 data categories for  $h = 1$ , and on average. For  $h = 4$  all three shrinkage estimators (LASSO, Elastic Net and the DFM) are doing equally well.

Hierarchical shrinkage priors based on the Uniform and inverse Gamma distributions are doing very poorly on average, although for some data categories they provide the smallest MAFE among all shrinkage estimators. Looking at the MAFE descriptives based on the total number of series, Student-t shrinkage is always doing better than Jeffreys shrinkage in lowering the median MAFE. Note however that for the Student-t prior, a single default choice of hyperparameters applied to all 129 series. Although this choice works well on average (median MAFEs), in some series it completely collapses. For example, there are cases where this estimate leads to MAFEs up to ten times higher (see the maximum MAFE based on total number of series in Table 3) than the benchmark model. On the other hand, Jeffrey's prior is not dependent on a choice of hyperparameters, and we can safely say that its shrinkage and forecasting performance is very unsatisfactory for the specific design of this study. Finally, the idea behind the fused LASSO, i.e. taking into account the correlation among consecutive predictors, does not help improve forecasting performance at all. In fact forecasts from this estimator are always dominated from the LASSO and the Elastic Net.

Once we turn to Tables 4 to 6 with the MSFE results based on the total number of series, it is obvious that the DFM is dominating all Bayesian shrinkage estimators at all three forecast horizons. Although the LASSO and the Elastic Net improve over the benchmark AR(2) forecasts,

they are still not as good as the DFM. Nevertheless, by looking at the individual data categories, the Elastic Net is the best in forecasting GDP and its components at horizons  $h = 1, 2$ , as well as the various Consumer Price Indexes at all forecast horizons. Note that this pattern was also true for the MAFE results in Tables 1 to 3. Therefore, summarizing the results in Tables 1 to 6, from a mean forecast error point of view the Elastic Net and the LASSO are the best Bayesian shrinkage estimators. However, these might not improve too much over principal component shrinkage using a factor model and the final result is dependent on the series being forecasted each time.

Table 7 gives a better view of the total performance of each shrinkage estimator. Hit rates are calculated based on MAFEs, MSFEs and predictive likelihoods. These are estimated as the proportion of times (among the 129 series) a specific shrinkage estimator had the lowest MAFE, the lowest MSFE and the highest average predictive likelihood (APL). The average predictive likelihood can be used to evaluate the whole predictive density of each regression model; see Geweke and Amisano (2010) and references therein. Although in Tables 1 to 6 we saw that based on the total number of series, the Elastic Net had exactly the same median MAFE and MSFE as the LASSO, Table 7 shows that the LASSO has better hit rates for all three measures. In terms of mean forecasting, the LASSO always does better in MSFE and MAFE hit rates by 8 to 15% compared to the Elastic Net. In terms of density forecasting, the LASSO improves even more the density forecasts from the Elastic Net (an average improvement of 25% at all forecast horizons). This is because parameter uncertainty feeds in the predictive likelihood evaluation. Thus the Elastic Net having two regularization parameters  $\lambda_1$  and  $\lambda_2$ , the uncertainty (posterior variance) about both parameters feeds in the density forecasts of  $y_{t+h}$ . The LASSO, having only one regularization parameter, i.e.  $\lambda_1 = \lambda$  and  $\lambda_2 = 0$ , has less forecast uncertainty/variance (given that for this specific case-study, forecasts of the *mean* coming from both estimators are more or less identical).

The similarity among the five shrinkage forecasts is assessed in Table 8. The lower triangular entries in this table show the correlation coefficients of all MSFEs for all 129 variables for horizon  $h = 1$ . The correlations among all shrinkage forecast errors is one, except for the Student-t forecasts which are less correlated to the other four shrinkage methods. This is simply because the other four hierarchical shrinkage priors (Jeffreys, LASSO, Fused LASSO and Elastic Net) are based on noninformative priors on the lower level of their hierarchies. Entries above the diagonal of Table 8 are the mean absolute difference between the row/column method RMSEs, averaged across series. The results confirm that the Student-t forecasts, which were the worst according to Tables 1-6, are the most distant from the forecasts generated from the other four priors. In contrast, the LASSO and Elastic Net forecasts have the smallest difference among any other method, something also confirmed by their equal forecasting performance shown in Tables 1-6.

No matter how correlated on average are the forecasts from the different shrinkage priors, we saw in Tables 1-6 that their differences were substantial across different data releases and across forecasts horizons. The LASSO and Elastic Net priors have a better ability to take into account the correlation patterns in the predictor variables, while the Fused LASSO is less good at this task (because of the very specific correlation pattern it has to find, i.e. penalize less/more consecutive predictors as a group). The Jeffreys and Student-t priors do not explicitly account for correlation in the predictor variables and, hence, their performance can be very risky sometimes,

with forecast errors which are multiples of those produced by the other three methods.

If correlation among the predictors is a crucial determinant of the performance of these algorithms, then a natural question to ask is what happens if we forecast with orthogonal predictors (the case that Stock and Watson, 2011, examine). Table 9 presents MAFE, MSFE and APL descriptive statistics based on all 129 series for  $h = 1$ , when the exogenous predictors are orthogonalized. For that reason the MATLAB function ORTH is used, which creates an orthonormal basis for the range of the matrix of exogenous predictors  $x_t$ , and which is based on simply taking the singular value decomposition of  $x_t$ . Consequently, this orthogonalization is like taking all possible principal components from the 128 exogenous predictors and then apply each of the five shrinkage algorithms to select the number of components to forecast with (while the rest are shrunk to zero). In fact, as seen on Table 9, orthogonalization of the data amounts to almost identical forecasting performance of the five Bayesian algorithms. Additionally their performance is equal to the best performing method, the LASSO, when using correlated predictors (compare the total MAFE and total MSFE results in Tables 1 and 4). This shows that orthogonalization is enough to guarantee that any of these shrinkage priors will always perform well in forecasting. However the reader should note that this happens due to the effect of the default, uninformative priors used in this paper. For informative choices on the regularization parameters the shrinkage penalty induced will - in general - be different among the five shrinkage priors (see the discussion in the following subsection).

### III.2 Forecasting one year ahead US GDP growth using the LASSO

The previous subsection focused on evaluating default semi-automatic shrinkage priors using 129 variables. In practical situations, the applied macroeconomist will most probably want to focus on a few variables of interest (like inflation, an output-gap or stock prices). Additionally, the previous subsection does not answer the question if other hyperparameter choices exist that could possibly make Bayesian shrinkage perform even better. Subsequently, here I focus on forecasting U.S. GDP using only the simple hierarchical LASSO prior, with various choices on the regularization parameter  $\lambda$ . In particular, following Del Mol et al. (2008) I use a regression model with an intercept and the 128 remaining variables as predictors (no own lags used).

The main difference with the previous subsection is that I compare four choices for  $\lambda$

1.  $\lambda^2 \sim \text{Gamma}(r, \delta)$  with  $r = \delta = 0.01$  (as in the benchmark case examined so far)
2.  $\lambda^2 \sim \text{Gamma}(r, \delta)$  with  $r = 1$  and  $\delta = 0.1$
3.  $\lambda^2 \sim \text{Gamma}(r, \delta)$  with  $r = 3$  and  $\delta = 1$
4.  $\lambda$  estimated by finding the maximum marginal likelihood (MML) using the Monte Carlo EM algorithm described in Park and Casella (2008)

Forecasts are generated for  $h = 4$  steps ahead, and MAFE and MSFE statistics relative to the random walk model are reported in Table 10. The benchmark prior is the best performing for US GDP and in fact forecasts are highly correlated with the DFM model. Using the full sample, estimates of the posterior median estimate of the regularization parameter  $\lambda$  in the four LASSO models are 87.2, 33.7, 10.8 and 10.6, respectively. The prior choice  $\lambda^2 \sim \text{Gamma}(3, 1)$  gives

posterior parameter estimates (and hence forecasts) identical to MML estimation of  $\lambda$ , and this actually occurs for a wide range of choices of  $r \geq 3$ .

As  $\lambda \rightarrow \infty$  all coefficients are penalized heavily, i.e.  $\beta_{\lambda \rightarrow \infty}^{LASSO} = 0$  which further implies that in the limit the dynamic regression model with many predictors reduces to  $y_{t+h} = \alpha z_t + \varepsilon_{t+h}$ . In this case, the scale invariant prior (benchmark case) provides the largest posterior estimate of  $\lambda$  which implies posterior estimates of  $\beta$  which are heavily penalized (but not exactly zero). As we allow informative priors (cases 2 and 3), more and more variables are left unrestricted and the results resemble the case of selection of regressors. For the third case, 14 coefficients are “sufficiently” different than zero, while the remaining 115 are very “low” (remember that the regressors are standardized, so it makes some sense to talk about “large” and “small” coefficients as being important or not). Nevertheless, one-year ahead forecasts of GDP growth are not improved when forecasting with these “14 predictors”, and hence the benchmark case which penalizes heavily all predictors performs better than using an informative prior on  $\lambda$ . This result is robust at other forecast horizons as well (results not presented here). The only difference is that as the forecast horizon increases (for  $h = 8$  for instance) more predictors are relevant for forecasting GDP, so that the  $\lambda^2 \sim \text{Gamma}(3, 1)$  prior leads to forecasts much closer to (but still dominated from) the choice  $\lambda^2 \sim \text{Gamma}(0.01, 0.01)$ .

## IV Concluding remarks

This paper has investigated the properties of Bayesian shrinkage using hierarchical priors. A general shrinkage representation is provided using Normal-Gamma distributions and five special cases of interest have been evaluated in forecasting using a large macroeconomic dataset. A default semi-automatic approach using noninformative, near-improper priors was given special attention in this paper, but also a sensitivity analysis with more informative priors has been carried out for forecasting US GDP.

The results suggest that Bayesian shrinkage can compete favorably with dynamic factor models, although it is not straightforward to say whether one method clearly dominates over the other. Both methods are efficient in reducing the dimension of large datasets and help achieve smaller forecast errors (especially for long-run forecasts), however extra care has to be taken when selecting a prior for Bayesian shrinkage. From an applied econometrician’s point of view (whether “frequentist” or “Bayesian”), the form of Bayesian shrinkage analyzed in this paper can be seen as a pragmatic tool useful for out-of-sample forecasting in the presence of many possible predictor variables (a typical every-day task for a researcher at the Fed, where thousands of series are available) or when time series are short (what is part of the life of a researcher in the ECB, with most Euro-Area macro series beginning around 1995). Subsequently, this paper argues that, similarly to the very popular Bayesian Model Averaging (BMA) and the empirical Bayes Minnesota prior for vector autoregressions, “formal” (i.e. hierarchical) Bayesian treatment of the shrinkage problem should also become a standard technique for handling modern medium to large amounts of information.



## References

- [1] Armagan, A. and Zaretzki, R. L. (2010). Model Selection via Adaptive Shrinkage with  $t$  Priors. *Computational Statistics* 25, 441-461.
- [2] Bai, J. and Ng, S. (2007). Boosting Diffusion Indexes. Unpublished manuscript, Columbia University.
- [3] De Mol, C., Giannone, D. and Reichlin, L. (2008). Forecasting Using a Large Number of Predictors: Is Bayesian Shrinkage a Valid Alternative to Principal Components? *Journal of Econometrics* 146, 318-328.
- [4] Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. (2004). Least Angle Regression. *The Annals of Statistics* 32, 407-451.
- [5] Efron, B. and Morris, C. (1975). Data Analysis Using Stein's Estimator and its Generalizations. *Journal of the American Statistical Association* 70, 311-319.
- [6] Fernandez, C., Ley, E. and Steel, M. F. J. (2001). Benchmark Priors for Bayesian Model Averaging. *Journal of Econometrics* 100, 381-427.
- [7] Gelman, A. (2006). Prior Distributions for Variance Parameters in Hierarchical Models. *Bayesian Analysis* 1, 515-533.
- [8] George, E and McCulloch, R. (1993). Variable Selection via Gibbs Sampling. *Journal of the American Statistical Association* 88, 881-889.
- [9] Geweke, J. (1993). Bayesian Treatment of the Independent Student- $t$  Linear Model. *Journal of Applied Econometrics* 8, 19-40.
- [10] Geweke, J. and Amisano, G. (2010). Comparing and evaluating Bayesian predictive distributions of asset returns. *International Journal of Forecasting* 26, 216-230.
- [11] Hobert, J. P and Casella, G. (1996). The Effect of Improper Priors on Gibbs Sampling in Hierarchical Mixed Models. *Journal of the American Statistical Association* 91, 1461-1473.
- [12] Hobert, J. P and Geyer, C. J. (1998). Geometric Ergodicity of Gibbs and Block Gibbs Samplers for a Hierarchical Random Effects Model. *Journal of Multivariate Analysis* 67, 414-430.
- [13] Inoue, A. and Kilian, L. (2008). How Useful Is Bagging in Forecasting Economic Time Series? A Case Study of U.S. Consumer Price Inflation. *Journal of the American Statistical Association* 103, 511-522.
- [14] Judge, G. G. and Bock, M.E. (1978). *Statistical Implications of Pre-Test and Stein Rule Estimators in Econometrics*. North-Holland, Amsterdam.
- [15] Koop, G. (2003). *Bayesian Econometrics*. Wiley, Chichester.

- [16] Koop, G. and Korobilis, D. (2009). Forecasting Inflation Using Dynamic Model Averaging, Working Paper Series 34-09, Rimini Centre for Economic Analysis.
- [17] Koop, G. and Potter, S. (2004). Forecasting in Dynamic Factor Models Using Bayesian Model Averaging. *The Econometrics Journal* 7, 550-565.
- [18] Kyung, M., Gill, J., Ghoshz, M. and Casella, G. (2010). Penalized Regression, Standard Errors, and Bayesian Lassos. *Bayesian Analysis* 5, 369-412.
- [19] Liang, F, Paulo, R., Molina, G., Clyde, M. A. and Berger, J. O. (2008). Mixtures of g-priors for Bayesian Variable Selection. *Journal of the American Statistical Association* 103, 410-423.
- [20] Litterman, R. (1979). Techniques of forecasting using vector autoregressions. Federal Reserve Bank of Minneapolis Working Paper 115.
- [21] Maruyama, Y. and George, E. I. (2010). *g*BF: A Fully Bayes Factor with a Generalized *g*-prior. Technical Report, University of Pennsylvania, available at <http://arxiv.org/abs/0801.4410>
- [22] Park, T. and Casella, G. (2008). The Bayesian Lasso. *Journal of the American Statistical Association* 103, 681-686.
- [23] Rosset, S. and Zhu, J. (2004), Discussion of “Least Angle Regression”, by B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani. *The Annals of Statistics* 32, 469–475
- [24] Stock, J. and Watson, M. (2011). Generalized Shrinkage Methods for Forecasting Using Many Predictors. Unpublished Manuscript, available at <http://www.princeton.edu/~mwatson/>.
- [25] Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society, Series B* 58, 267-288.
- [26] Tibshirani, R., Saunders, M., Rosset, S., Zhu, J. and Knight, K. (2005). Sparsity and Smoothness via the Fused Lasso. *Journal of the Royal Statistical Society, Series B* 67, 91-108.
- [27] Zellner, A. (1986). On Assessing Prior Distributions and Bayesian Regression Analysis with g-prior Distributions. in: P. Goel and A. Zellner (Eds.) *Bayesian Inference and Decision Techniques* (North-Holland, Amsterdam).
- [28] Zou, H. (2006). The Adaptive Lasso and Its Oracle Properties. *Journal of the American Statistical Association* 101, 1418-1429.
- [29] Zou, H. and Hastie, T. (2005). Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society, Series B* 67, 301-320.

## Appendices

### A. Data and transformations

All series were downloaded from St. Louis' FRED database in December 2010 and cover the quarters Q1:1959 to Q2:2010. All series were seasonally adjusted: either taken adjusted from FRED or by applying to the unadjusted series a quarterly X11 filter based on an AR(4) model (after testing for seasonality). Some series in the database were observed only on a monthly basis and quarterly values were computed by averaging the monthly values over the quarter (as opposed to keeping the mid-month of the quarter). All variables are transformed to be approximately stationary and the transformation codes for each variable appear in the column 'T' on the table below.

In particular, if  $w_{i,t}$  is the original un-transformed series in levels, when the series is used as a predictor (R.H.S. of equation (1)) the transformation codes are: 1 - no transformation (levels),  $x_{i,t} = w_{i,t}$ ; 2 - first difference,  $x_{i,t} = w_{i,t} - w_{i,t-1}$ ; 3- second difference,  $x_{i,t} = \Delta w_{i,t} - \Delta w_{i,t-1}$  4 - logarithm,  $x_{i,t} = \log w_{i,t}$ ; 5 - first difference of logarithm,  $x_{i,t} = \log w_{i,t} - \log w_{i,t-1}$ ; 6 - second difference of logarithm,  $x_{i,t} = \Delta \log w_{i,t} - \Delta \log w_{i,t-1}$ .

When the series is used as the variable to be predicted (L.H.S. of equation (1)) the transformation codes are: 1 - no transformation (levels),  $y_{i,t+h} = w_{i,t+h}$ ; 2 - first difference,  $y_{i,t+h} = w_{i,t+h} - w_{i,t}$ ; 3- second difference,  $y_{i,t+h} = \frac{1}{h} \Delta^h w_{i,t+h} - \Delta w_{i,t}$  4 - logarithm,  $y_{i,t+h} = \log w_{i,t+h}$ ; 5 - first difference of logarithm,  $y_{i,t+h} = \log w_{i,t+h} - \log w_{i,t}$ ; 6 - second difference of logarithm,  $y_{i,t+h} = \frac{1}{h} \Delta^h \log w_{i,t+h} - \Delta \log w_{i,t}$ . In the transformations above, I define  $\Delta w_t = w_t - w_{t-1}$  and  $\Delta^h w_{t+h} = w_{t+h} - w_t$ .

From the 129 series, 14 are higher level aggregates and do not add information when extracting principal components. These series are indicated with a 0 in column 'F' of the table below, and only the rest 115 series are used for estimating factors.

**Table A.1: Description of series**

| No | Series ID | T | F | Title   |
|----|-----------|---|---|---|
| 1  | GDPC96    | 5 | 1 | Real Gross Domestic Product, 3 Decimal                      |
| 2  | GDPDEF    | 5 | 1 | Gross Domestic Product: Implicit Price Deflator             |
| 3  | PCECC96   | 5 | 1 | Real Personal Consumption Expenditures                      |
| 4  | PCECTPI   | 5 | 1 | Personal Consumption Expenditures: Chain-type Price Index   |
| 5  | GPDI96    | 5 | 1 | Real Gross Private Domestic Investment, 3 Decimal           |
| 6  | IMPGSC96  | 5 | 1 | Real Imports of Goods & Services, 3 Decimal                 |
| 7  | EXPGSC96  | 5 | 1 | Real Exports of Goods & Services, 3 Decimal                 |
| 8  | CBIC96    | 1 | 1 | Real Change in Private Inventories                          |
| 9  | FINSLC96  | 5 | 1 | Real Final Sales of Domestic Product                        |
| 10 | GSAVE     | 5 | 1 | Gross Saving  |
| 11 | GCEC96    | 5 | 1 | Real Government Consumption Expenditures & Gross Investment |
| 12 | SLEXPND   | 6 | 1 | State & Local Government Current Expenditures               |
| 13 | SLINV     | 6 | 1 | State & Local Government Gross Investment                   |
| 14 | DPIC96    | 6 | 1 | Real Disposable Personal Income                             |
| 15 | PINCOME   | 6 | 1 | Personal Income   |
| 16 | PSAVE     | 5 | 1 | Personal Saving   |
| 17 | PRFI      | 6 | 1 | Private Residential Fixed Investment                        |

| No | Series ID | T | F | Title  |
|----|-----------|---|---|--|
| 18 | PNFI      | 6 | 1 | Private Nonresidential Fixed Investment                          |
| 19 | PCDG      | 5 | 1 | Personal Consumption Expenditures: Durable Goods                 |
| 20 | PCND      | 5 | 1 | Personal Consumption Expenditures: Nondurable Goods              |
| 21 | PCESV     | 5 | 1 | Personal Consumption Expenditures: Services                      |
| 22 | GPDICTPI  | 6 | 1 | Gross Private Domestic Investment: Chain-type Price Index        |
| 23 | WASCUR    | 6 | 1 | Compensation of Employees: Wages & Salary Accruals               |
| 24 | DIVIDEND  | 6 | 1 | Net Corporate Dividends  |
| 25 | CP        | 6 | 1 | Corporate Profits After Tax                                      |
| 26 | CCFC      | 6 | 1 | Corporate: Consumption of Fixed Capital                          |
| 27 | HOUST     | 4 | 0 | Housing Starts: Total: New Privately Owned Housing Units Started |
| 28 | HOUST1F   | 4 | 1 | Privately Owned Housing Starts: 1-Unit Structures                |
| 29 | HOUST5F   | 4 | 1 | Privately Owned Housing Starts: 5-Unit Structures or More        |
| 30 | HOUSTW    | 4 | 1 | Housing Starts in West Census Region                             |
| 31 | HOUSTMW   | 4 | 1 | Housing Starts in Midwest Census Region                          |
| 32 | HOUSTS    | 4 | 1 | Housing Starts in South Census Region                            |
| 33 | HOUSTNE   | 4 | 1 | Housing Starts in Northeast Census Region                        |
| 34 | INDPRO    | 5 | 0 | Industrial Production Index                                      |
| 35 | IPCONGD   | 5 | 1 | Industrial Production: Consumer Goods                            |
| 36 | IPDCONGD  | 5 | 1 | Industrial Production: Durable Consumer Goods                    |
| 37 | IPNCONGD  | 5 | 1 | Industrial Production: Nondurable Consumer Goods                 |
| 38 | IPMAT     | 5 | 1 | Industrial Production: Materials                                 |
| 39 | IPDMAT    | 5 | 1 | Industrial Production: Durable Materials                         |
| 40 | IPNMAT    | 5 | 1 | Industrial Production: Nondurable Materials                      |
| 41 | IPBUSEQ   | 5 | 1 | Industrial Production: Business Equipment                        |
| 42 | IPFINAL   | 5 | 1 | Industrial Production: Final Products (Market Group)             |
| 43 | UTL11     | 1 | 1 | Capacity Utilization: Manufacturing                              |
| 44 | UEMPLT5   | 5 | 1 | Civilians Unemployed - Less Than 5 Weeks                         |
| 45 | UEMP5TO14 | 5 | 1 | Civilians Unemployed for 5-14 Weeks                              |
| 46 | UEMP15T26 | 5 | 1 | Civilians Unemployed for 15-26 Weeks                             |
| 47 | UEMP27OV  | 5 | 1 | Civilians Unemployed for 27 Weeks and Over                       |
| 48 | UNRATE    | 2 | 1 | Civilian Unemployment Rate                                       |
| 49 | PAYEMS    | 5 | 0 | Total Nonfarm Payrolls: All Employees                            |
| 50 | NDMANEMP  | 5 | 1 | All Employees: Nondurable Goods Manufacturing                    |
| 51 | DMANEMP   | 5 | 1 | All Employees: Durable Goods Manufacturing                       |
| 52 | USCONS    | 5 | 1 | All Employees: Construction                                      |
| 53 | USGOOD    | 5 | 0 | All Employees: Goods-Producing Industries                        |
| 54 | USFIRE    | 5 | 1 | All Employees: Financial Activities                              |
| 55 | USWTRADE  | 5 | 1 | All Employees: Wholesale Trade                                   |
| 56 | USTPU     | 5 | 1 | All Employees: Trade, Transportation & Utilities                 |
| 57 | USTRADE   | 5 | 1 | All Employees: Retail Trade                                      |
| 58 | USMINE    | 5 | 1 | All Employees: Natural Resources & Mining                        |
| 59 | USPBS     | 5 | 1 | All Employees: Professional & Business Services                  |
| 60 | USLAH     | 5 | 1 | All Employees: Leisure & Hospitality                             |
| 61 | USINFO    | 5 | 1 | All Employees: Information Services                              |
| 62 | USEHS     | 5 | 1 | All Employees: Education & Health Services                       |
| 63 | SRVPRD    | 5 | 1 | All Employees: Service-Providing Industries                      |
| 64 | USPRIV    | 5 | 0 | All Employees: Total Private Industries                          |
| 65 | USGOVT    | 5 | 1 | All Employees: Government  |
| 66 | AHEMAN    | 6 | 1 | Average Hourly Earnings: Manufacturing                           |

| No  | Series ID | T | F | Title   |
|-----|-----------|---|---|---|
| 67  | AHECONS   | 6 | 1 | Average Hourly Earnings: Construction                               |
| 68  | AWHMAN    | 5 | 1 | Average Weekly Hours of Production: Manufacturing                   |
| 69  | AWOTMAN   | 5 | 1 | Average Weekly Hours: Overtime: Manufacturing                       |
| 70  | EMRATIO   | 5 | 1 | Civilian Employment-Population Ratio                                |
| 71  | CIVPART   | 5 | 1 | Civilian Participation Rate   |
| 72  | OPHPBS    | 5 | 1 | Business Sector: Output Per Hour of All Persons                     |
| 73  | ULCNFB    | 5 | 1 | Nonfarm Business Sector: Unit Labor Cost                            |
| 74  | BUSLOANS  | 6 | 1 | Commercial and Industrial Loans at All Commercial Banks             |
| 75  | REALLN    | 6 | 1 | Real Estate Loans at All Commercial Banks                           |
| 76  | CONSUMER  | 5 | 1 | Consumer (Individual) Loans at All Commercial Banks                 |
| 77  | INVEST    | 5 | 0 | Total Investments at All Commercial Banks                           |
| 78  | LOANS     | 6 | 0 | Total Loans and Leases at Commercial Banks                          |
| 79  | MPRIME    | 2 | 1 | Bank Prime Loan Rate  |
| 80  | GS1       | 2 | 1 | 1-Year Treasury Constant Maturity Rate                              |
| 81  | GS3       | 2 | 1 | 3-Year Treasury Constant Maturity Rate                              |
| 82  | GS5       | 2 | 1 | 5-Year Treasury Constant Maturity Rate                              |
| 83  | GS10      | 2 | 1 | 10-Year Treasury Constant Maturity Rate                             |
| 84  | FEDFUNDS  | 2 | 1 | Effective Federal Funds Rate  |
| 85  | TB3MS     | 2 | 1 | 3-Month Treasury Bill: Secondary Market Rate                        |
| 86  | TB6MS     | 2 | 1 | 6-Month Treasury Bill: Secondary Market Rate                        |
| 87  | AAA       | 2 | 1 | Moody's Seasoned Aaa Corporate Bond Yield                           |
| 88  | BAA       | 2 | 1 | Moody's Seasoned Baa Corporate Bond Yield                           |
| 89  | M1SL      | 6 | 1 | M1 Money Stock  |
| 90  | M2SL      | 6 | 1 | M2 Money Stock  |
| 91  | CURRSL    | 6 | 1 | Currency Component of M1  |
| 92  | DEMDEPSL  | 6 | 1 | Demand Deposits at Commercial Banks                                 |
| 93  | SAVINGSL  | 6 | 1 | Savings Deposits - Total  |
| 94  | TCDSL     | 6 | 0 | Total Checkable Deposits  |
| 95  | TVCKSSL   | 6 | 1 | Travelers Checks Outstanding  |
| 96  | CURRCIR   | 6 | 1 | Currency in Circulation   |
| 97  | MZMSL     | 6 | 1 | MZM Money Stock   |
| 98  | M1V       | 5 | 1 | Velocity of M1 Money Stock  |
| 99  | M2V       | 5 | 1 | Velocity of M2 Money Stock  |
| 100 | NONREVSL  | 6 | 0 | Total Nonrevolving Credit Outstanding                               |
| 101 | TOTALSL   | 6 | 0 | Total Consumer Credit Outstanding                                   |
| 102 | CPIAUCSL  | 6 | 0 | Consumer Price Index for All Urban Consumers: All Items             |
| 103 | CPILEGSL  | 6 | 0 | Consumer Price Index for All Urban Consumers: All Items Less Energy |
| 104 | CPIULFSL  | 6 | 0 | Consumer Price Index for All Urban Consumers: All Items Less Food   |
| 105 | CPIENGSL  | 6 | 1 | Consumer Price Index for All Urban Consumers: Energy                |
| 106 | CPIUFDSL  | 6 | 1 | Consumer Price Index for All Urban Consumers: Food                  |
| 107 | CPIAPPSL  | 6 | 1 | Consumer Price Index for All Urban Consumers: Apparel               |
| 108 | CPIMEDSL  | 6 | 1 | Consumer Price Index for All Urban Consumers: Medical Care          |
| 109 | CPITRNSL  | 6 | 1 | Consumer Price Index for All Urban Consumers: Transportation        |
| 110 | PPIACO    | 6 | 0 | Producer Price Index: All Commodities                               |
| 111 | PPIFCG    | 6 | 1 | Producer Price Index: Finished Consumer Goods                       |
| 112 | PPIFCF    | 6 | 1 | Producer Price Index: Finished Consumer Foods                       |
| 113 | PFCGEF    | 6 | 1 | Producer Price Index: Finished Consumer Goods Excluding Foods       |
| 114 | PPIFGS    | 6 | 1 | Producer Price Index: Finished Goods                                |

| No  | Series ID | T | F | Title   |
|-----|-----------|---|---|---|
| 115 | PPICRM    | 6 | 1 | Producer Price Index: Crude Materials for Further Processing        |
| 116 | PPICPE    | 6 | 1 | Producer Price Index Finished Goods: Capital Equipment              |
| 117 | PPIITM    | 6 | 1 | Producer Price Index: Intermediate Materials: Supplies & Components |
| 118 | SP500     | 5 | 1 | S&P 500 Index   |
| 119 | EXUSUK    | 5 | 1 | U.S. / U.K Foreign Exchange Rate                                    |
| 120 | EXSZUS    | 5 | 1 | Switzerland / U.S. Foreign Exchange Rate                            |
| 121 | EXJPUS    | 5 | 1 | Japan / U.S. Foreign Exchange Rate                                  |
| 122 | EXCAUS    | 5 | 1 | Canada / U.S. Foreign Exchange Rate                                 |
| 123 | PMI       | 1 | 1 | ISM Manufacturing: PMI Composite Index                              |
| 124 | NAPMNOI   | 1 | 1 | ISM Manufacturing: New Orders Index                                 |
| 125 | NAPMII    | 1 | 1 | ISM Manufacturing: Inventories Index                                |
| 126 | NAPMEI    | 1 | 1 | ISM Manufacturing: Employment Index                                 |
| 127 | NAPMPRI   | 1 | 1 | ISM Manufacturing: Prices Index                                     |
| 128 | NAPMPI    | 1 | 1 | ISM Manufacturing: Production Index                                 |
| 129 | NAPMSDI   | 1 | 1 | ISM Manufacturing: Supplier Deliveries Index                        |

**Table A.2: Categories of data series based on statistical releases**

| Group | Release   | Number of series |
|-------|---|------------------|
| 1     | Gross Domestic Product  | 26               |
| 2     | New Residential Construction  | 7                |
| 3     | G.17 Industrial Production and Capacity Utilization                 | 10               |
| 4     | The Employment Situation  | 28               |
| 5     | Productivity and Costs  | 2                |
| 6     | H.8 Assets and Liabilities of Commercial Banks in the United States | 5                |
| 7     | H.15 Selected Interest Rates  | 10               |
| 8     | H.6 Money Stock Measures  | 7                |
| 9     | H.4.1 Factors Affecting Reserve Balances                            | 1                |
| 10    | Money Zero Maturity (MZM)   | 1                |
| 11    | Money Velocity  | 2                |
| 12    | G.19 Consumer Credit  | 2                |
| 13    | Consumer Price Index  | 8                |
| 14    | Producer Price Index  | 8                |
| 15    | Standard & Poors  | 1                |
| 16    | G.5 Foreign Exchange Rates  | 4                |
| 17    | Manufacturing ISM Report on Business                                | 7                |

## B. Bayesian hierarchical shrinkage priors

Note that I denote the inverse Gaussian distribution with parameters  $c, d$  as  $IG(c, d)$ , while the inverse Gamma with parameters  $a, b$  is denoted as  $iGamma(a, b)$ . A variable coming from the inverse Gamma distribution is the reciprocal of a variable distributed as gamma, while the same is *not* true for the inverse Gaussian variate (i.e. if  $z \sim IG(c, d)$ , then  $(z^{-1}) \approx N(c, d)$ ). There are many parametrizations of the Gamma distribution, and the one I am using in this article is

$$Gamma(a, b) \equiv f(y; a, b) = C_G y^{a-1} b^a e^{-by}$$

for  $a, b$  non-negative, real numbers, where  $C_G = \Gamma(a) = (a-1)!$  is the gamma function.

The parameters on the unrestricted variables  $z_t$  are integrated out with the noninformative prior  $\pi(\alpha) \propto 1$ , leading to a conditional posterior

$$\alpha | \beta, \sigma^2, data \sim N_q \left( (Z'Z)^{-1} Z' \tilde{y}^\beta, \sigma^2 (Z'Z)^{-1} \right) \quad (B.1)$$

with  $Z = (z'_1, \dots, z'_T)'$ . That is, in the formulas of the conditional posteriors below, we need to add in each and every case of hierarchical prior the sampling step in equation (B.1) above. For notational convenience, in (B.1) and in the conditional posteriors below some or all of the quantities  $\tilde{y}^\beta$ ,  $\tilde{y}^a$  and  $\Psi$  show up, which are defined as  $\tilde{y}^\beta = y - X\beta$ ,  $\tilde{y}^a = y - Z\alpha$  and  $\Psi = (y - Z\alpha - X\beta)'(y - Z\alpha - X\beta)$ , respectively. *Finally, in the formulas for the conditional posteriors we have to condition on the data matrices  $(y, Z, X)$ , but this is omitted for notational simplicity (to keep the formulas more compact).*

### B.1 Adaptive shrinkage Jeffrey's prior

The priors are defined using the following hierarchy

$$\begin{aligned} \pi(\beta | \tau_1^2, \dots, \tau_p^2) &\sim N_p(0, V) \\ \pi(\tau_j^2) &\sim 1/\tau_j^2, \text{ for } j = 1, \dots, p \end{aligned}$$

where  $V = diag\{\tau_1^2, \dots, \tau_p^2\}$ . The posteriors of  $\beta$  and  $\tau_j^2$  can be obtained by sampling recursively from (B.1) and the full conditionals

$$\beta | a, \sigma^2, \{\tau_j^2\}_{j=1}^p \sim N_p \left( (X'X + \sigma^2 V^{-1})^{-1} X' \tilde{y}^a, \sigma^2 (X'X + \sigma^2 V^{-1})^{-1} \right) \quad (B.2a)$$

$$\frac{1}{\tau_j^2} | a, \beta_j, \sigma^2 \sim Gamma \left( \frac{1}{2}, \frac{\beta_j^2}{2} \right), \text{ for } j = 1, \dots, p \quad (B.2b)$$

$$\sigma^2 | a, \beta, \{\tau_j^2\}_{j=1}^p \sim iGamma \left( \frac{T}{2}, \frac{1}{2} \Psi \right) \quad (B.2c)$$

## B.2 Adaptive shrinkage t-prior

The prior for this case is of the hierarchical form

$$\begin{aligned}\pi(\beta|\sigma^2, \tau_1^2, \dots, \tau_p^2) &\sim N_p(0, \sigma^2 V) \\ \pi(\tau_j^2) &\sim iGamma(\rho, \xi), \text{ for } j = 1, \dots, p\end{aligned}$$

where  $V = diag\{\tau_1^2, \dots, \tau_p^2\}$ . Draws from the posterior can be obtained by sampling recursively from (B.1) and the full conditionals

$$\beta|a, \sigma^2, \{\tau_j^2\}_{j=1}^p \sim N_p\left((X'X + \sigma^2 V^{-1})^{-1} X' \tilde{y}^a, \sigma^2 (X'X + \sigma^2 V^{-1})^{-1}\right) \quad (\text{B.3a})$$

$$\frac{1}{\tau_j^2}|a, \beta_j, \sigma^2 \sim Gamma\left(\rho + \frac{1}{2}, \frac{\beta_j^2}{2} + \xi\right), \text{ for } j = 1, \dots, p \quad (\text{B.3b})$$

$$\sigma^2|a, \beta, \{\tau_j^2\}_{j=1}^p \sim iGamma\left(\frac{T}{2}, \frac{1}{2}\Psi\right) \quad (\text{B.3c})$$

## B.3 Hierarchical Lasso

The full hierarchical representation of the LASSO prior is

$$\pi(\beta|\sigma^2, \tau_1^2, \dots, \tau_p^2) \sim N_p(0, \sigma^2 V) \quad (\text{B.4a})$$

$$\pi(\tau_j^2) \sim Exponential\left(\frac{\lambda^2}{2}\right), \text{ for } j = 1, \dots, p \quad (\text{B.4b})$$

$$\pi(\lambda^2) \sim Gamma(r, \delta) \quad (\text{B.4c})$$

where  $V = diag\{\tau_1^2, \dots, \tau_p^2\}$ .

Given these priors, the posterior can be obtained by sampling recursively from (B.1) and the full conditionals

$$\beta|a, \sigma^2, \{\tau_j^2\}_{j=1}^p \sim N_p\left((X'X + V^{-1})^{-1} X' \tilde{y}^a, \sigma^2 (X'X + V^{-1})^{-1}\right) \quad (\text{B.5a})$$

$$\frac{1}{\tau_j^2}|a, \beta, \sigma^2 \sim IG\left(\sqrt{\frac{\lambda^2 \sigma^2}{\beta_j^2}}, \lambda^2\right), \text{ for } j = 1, \dots, p \quad (\text{B.5b})$$

$$\lambda^2|a, \beta, \sigma^2, \{\tau_j^2\}_{j=1}^p \sim Gamma\left(p + r, \frac{1}{2} \sum_{j=1}^p \tau_j^2 + \delta\right) \quad (\text{B.5c})$$

$$\sigma^2|a, \beta, \{\tau_j^2\}_{j=1}^p \sim iGamma\left(\frac{T-1}{2} + \frac{p}{2}, \frac{1}{2}\Psi + \frac{1}{2}\beta' V^{-1} \beta\right) \quad (\text{B.5d})$$



## B.4 Hierarchical Fused Lasso

The hierarchical representation of the Fused Lasso prior is

$$\pi(\beta|\sigma^2, \tau_1^2, \dots, \tau_p^2) \sim N_p(0, \sigma^2 V) \quad (\text{B.6a})$$

$$\pi(\tau_j^2|\lambda_1) \sim \text{Exponential}\left(\frac{\lambda_1^2}{2}\right), \text{ for } j = 1, \dots, p \quad (\text{B.6b})$$

$$\pi(\omega_j^2|\lambda_2) \sim \text{Exponential}\left(\frac{\lambda_2^2}{2}\right), \text{ for } j = 1, \dots, p-1 \quad (\text{B.6c})$$

$$\pi(\lambda_1^2) \sim \text{Gamma}(r_1, \delta_1) \quad (\text{B.6d})$$

$$\pi(\lambda_2^2) \sim \text{Gamma}(r_2, \delta_2) \quad (\text{B.6e})$$

where in this case  $V$  is the tridiagonal matrix

$$V = \sigma^2 \times \begin{bmatrix} (\tau_1^2 + \omega_0^2 + \omega_1^2) & -\omega_1^2 & 0 & \dots & 0 \\ -\omega_1^2 & (\tau_2^2 + \omega_1^2 + \omega_2^2) & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & -\omega_{p-2}^2 & 0 \\ \vdots & \ddots & -\omega_{p-2}^2 & (\tau_{p-1}^2 + \omega_{p-2}^2 + \omega_{p-1}^2) & -\omega_{p-1}^2 \\ 0 & \dots & 0 & -\omega_{p-1}^2 & (\tau_p^2 + \omega_{p-1}^2 + \omega_p^2) \end{bmatrix}.$$

Given these priors, the posteriors can be obtained by sampling recursively from (B.1) and the full conditionals

$$\beta|a, \sigma^2, \{\tau_j^2\}_{j=1}^p, \{\omega_j^2\}_{j=1}^{p-1} \sim N_p\left((X'X + V_{FL}^{-1})^{-1} X' \tilde{y}^a, \sigma^2 (X'X + V^{-1})^{-1}\right) \quad (\text{B.7a})$$

$$\frac{1}{\tau_j^2}|a, \beta, \{\omega_j^2\}_{j=1}^{p-1}, \sigma^2 \sim IG\left(\sqrt{\frac{\lambda_1^2 \sigma^2}{\beta_j^2}}, \lambda^2\right), \text{ for } j = 1, \dots, p \quad (\text{B.7b})$$

$$\frac{1}{\omega_j^2}|a, \beta, \{\tau_j^2\}_{j=1}^p, \sigma^2 \sim IG\left(\sqrt{\frac{\lambda_2^2 \sigma^2}{(\beta_{j+1} - \beta_j)^2}}, \lambda^2\right), \text{ for } j = 1, \dots, p-1 \quad (\text{B.7c})$$

$$\lambda_1^2|a, \beta, \sigma^2, \{\tau_j^2\}_{j=1}^p, \{\omega_j^2\}_{j=1}^{p-1} \sim \text{Gamma}\left(p + r_1, \frac{1}{2} \sum_{j=1}^p \tau_j^2 + \delta_1\right) \quad (\text{B.7d})$$

$$\lambda_2^2|a, \beta, \sigma^2, \{\tau_j^2\}_{j=1}^p, \{\omega_j^2\}_{j=1}^{p-1} \sim \text{Gamma}\left(p - 1 + r_2, \frac{1}{2} \sum_{j=1}^{p-1} \omega_j^2 + \delta_2\right) \quad (\text{B.7e})$$

$$\sigma^2|a, \beta, \{\tau_j^2\}_{j=1}^p \sim i\text{Gamma}\left(\frac{T-1}{2} + \frac{p}{2}, \frac{1}{2} \Psi + \frac{1}{2} \beta' V^{-1} \beta\right) \quad (\text{B.7f})$$

## B.5 Hierarchical Elastic Net

For a covariance  $V = \sigma^2 \times \text{diag}\left\{(\tau_1^{-2} + \lambda_2)^{-1}, \dots, (\tau_p^{-2} + \lambda_2)^{-1}\right\}$  matrix the hierarchical elastic

net prior is

$$\begin{aligned}
\pi(\beta|\sigma^2, \tau_1^2, \dots, \tau_p^2) &\sim N_p(0, V) \\
\pi(\tau_j^2|\lambda_1^2) &\sim \text{Exponential}\left(\frac{\lambda_1^2}{2}\right), \text{ for } j = 1, \dots, p \\
\pi(\lambda_1^2) &\sim \text{Gamma}(r_1, \delta_1) \\
\pi(\lambda_2^2) &\sim \text{Gamma}(r_2, \delta_2)
\end{aligned}$$

Given these priors, the posterior can be obtained by sampling recursively from (B.1) and the full conditionals

$$\beta|a, \sigma^2, \{\tau_j^2\}_{j=1}^p \sim N_p\left((X'X + V^{-1})^{-1} X' \hat{y}^a, \sigma^2 (X'X + V^{-1})^{-1}\right) \quad (\text{B.8a})$$

$$\frac{1}{\tau_j^2}|a, \beta, \sigma^2 \sim IG\left(\sqrt{\frac{\lambda_1^2 \sigma^2}{\beta_j^2}}, \lambda_1^2\right), \text{ for } j = 1, \dots, p \quad (\text{B.8b})$$

$$\lambda_1^2|a, \beta, \sigma^2, \{\tau_j^2\}_{j=1}^p \sim \text{Gamma}\left(p + r_1, \frac{1}{2} \sum_{j=1}^p \tau_j^2 + \delta_1\right) \quad (\text{B.8c})$$

$$\lambda_2^2|a, \beta, \sigma^2, \{\tau_j^2\}_{j=1}^p \sim \text{Gamma}\left(\frac{p}{2} + r_2, \frac{1}{2\sigma^2} \sum_{j=1}^p \beta_j^2 + \delta_2\right) \quad (\text{B.8d})$$

$$\sigma^2|a, \beta, \{\tau_j^2\}_{j=1}^p \sim i\text{Gamma}\left(\frac{T-1}{2} + \frac{p}{2}, \frac{1}{2} \Psi + \frac{1}{2} \beta' V^{-1} \beta\right) \quad (\text{B.8e})$$

## C. Results: Tables

Table 1: MAFE results for the five Bayes shrinkage estimators and the DFM,  $h = 1$

|  | Jef          | St-t  | LASSO        | Fused<br>LASSO | Elastic<br>Net | DFM          |
|--|--------------|-------|--------------|----------------|----------------|--------------|
| <i>Median MAFE's based on statistical releases</i>       |              |       |              |                |                |              |
| GDP and components                                       | 1.064        | 1.001 | 0.987        | 1.016          | <b>0.986</b>   | 1.030        |
| Housing  | 1.255        | 1.020 | 1.039        | 1.019          | 1.040          | <b>0.924</b> |
| IP   | 1.016        | 1.027 | 0.988        | 1.013          | 0.988          | <b>0.928</b> |
| Employment situation                                     | 1.058        | 1.024 | 1.012        | 1.033          | 1.012          | <b>0.976</b> |
| Productivity/Costs                                       | 1.018        | 1.169 | 0.986        | 1.043          | 0.987          | <b>0.959</b> |
| Assets/Liabilities of banks                              | 0.966        | 0.973 | 0.970        | 0.973          | 0.971          | <b>0.957</b> |
| Interest rates   | 1.056        | 0.970 | 0.967        | 0.991          | <b>0.966</b>   | 1.066        |
| Money stock  | 0.917        | 0.910 | 0.905        | 0.908          | <b>0.904</b>   | 1.050        |
| Currecny in circulation                                  | 0.677        | 0.680 | 0.678        | <b>0.671</b>   | 0.677          | 0.678        |
| MZM  | 0.910        | 0.894 | 0.896        | 0.900          | <b>0.893</b>   | 1.294        |
| Money velocity   | 0.994        | 0.976 | 0.981        | 0.990          | <b>0.979</b>   | 1.092        |
| Consumer Credit  | 1.001        | 0.985 | 0.985        | 0.998          | <b>0.980</b>   | 1.043        |
| CPI  | 1.006        | 0.914 | 0.916        | 0.914          | <b>0.909</b>   | 0.963        |
| PPI  | <b>0.913</b> | 0.917 | 0.916        | 0.918          | 0.919          | 0.931        |
| Stock prices   | <b>1.003</b> | 1.008 | 1.006        | 1.004          | 1.005          | 1.072        |
| Exchange rates   | 0.999        | 1.004 | 1.003        | <b>0.997</b>   | 1.003          | 0.999        |
| ISM surveys  | 1.108        | 1.276 | 1.031        | 1.038          | 1.028          | <b>0.987</b> |
| <i>MAFE descriptives based on total number of series</i> |              |       |              |                |                |              |
| median   | 1.017        | 0.998 | <b>0.987</b> | 1.004          | <b>0.987</b>   | 0.993        |
| variance   | 0.045        | 0.509 | 0.008        | 0.015          | 0.007          | 0.015        |
| min  | 0.677        | 0.680 | 0.678        | 0.671          | 0.677          | 0.678        |
| max  | 1.857        | 2.342 | 1.435        | 1.645          | 1.286          | 1.582        |

Note: Entries are MAFE-based statistics relative to the MAFE of an AR(2) model.

Table 2: MAFE results for the five Bayes shrinkage estimators and the DFM,  $h = 2$ 

|  | Jef          | St-t         | LASSO        | Fused<br>LASSO | Elastic<br>Net | DFM          |
|--|--------------|--------------|--------------|----------------|----------------|--------------|
| <i>Median MAFE's based on statistical releases</i>       |              |              |              |                |                |              |
| GDP and components                                       | 1.100        | 1.025        | <b>0.988</b> | 1.025          | <b>0.988</b>   | 1.005        |
| Housing  | 1.208        | 1.045        | 1.040        | 1.035          | 1.043          | <b>1.017</b> |
| IP   | 1.050        | 1.086        | <b>0.987</b> | 1.026          | 0.989          | 0.989        |
| Employment situation                                     | 1.089        | 1.029        | 1.017        | 1.039          | 1.016          | <b>0.982</b> |
| Productivity/Costs                                       | 1.160        | 1.386        | 1.081        | 1.184          | 1.082          | <b>0.942</b> |
| Assets/Liabilities of banks                              | <b>0.967</b> | 1.072        | 0.973        | 0.977          | 0.969          | 0.968        |
| Interest rates   | 1.294        | 0.998        | <b>0.986</b> | 1.082          | 0.987          | 1.000        |
| Money stock  | <b>0.929</b> | 0.933        | 0.932        | 0.930          | 0.931          | 1.002        |
| Currecny in circulation                                  | 0.737        | 0.736        | 0.737        | <b>0.733</b>   | 0.735          | 0.746        |
| MZM  | 0.855        | 0.855        | <b>0.845</b> | 0.847          | 0.849          | 1.112        |
| Money velocity   | 1.000        | <b>0.995</b> | 0.996        | 1.002          | 0.998          | 1.029        |
| Consumer Credit  | <b>0.990</b> | 0.999        | 0.996        | <b>0.990</b>   | 0.996          | 1.010        |
| CPI  | 1.080        | 0.962        | 0.955        | 0.979          | <b>0.954</b>   | 0.994        |
| PPI  | 0.958        | 0.942        | <b>0.941</b> | 0.957          | 0.945          | 0.975        |
| Stock prices   | <b>0.987</b> | 0.995        | 0.997        | 0.992          | 0.996          | 1.049        |
| Exchange rates   | 1.026        | 1.007        | 1.008        | <b>1.005</b>   | 1.009          | 1.030        |
| ISM surveys  | 1.041        | 1.415        | <b>1.021</b> | 1.057          | <b>1.021</b>   | 1.050        |
| <i>MAFE descriptives based on total number of series</i> |              |              |              |                |                |              |
| median   | 1.038        | 1.015        | <b>0.990</b> | 1.019          | 0.991          | 0.999        |
| variance   | 0.296        | 0.625        | 0.009        | 0.029          | 0.008          | 0.007        |
| min  | 0.491        | 0.494        | 0.487        | 0.495          | 0.482          | 0.488        |
| max  | 3.230        | 4.536        | 1.458        | 1.931          | 1.289          | 1.222        |

Note: Entries are MAFE-based statistics relative to the MAFE of an AR(2) model.

Table 3: MAFE results for the five Bayes shrinkage estimators and the DFM,  $h = 4$

|  | Jef          | St-t         | LASSO        | Fused<br>LASSO | Elastic<br>Net | DFM          |
|--|--------------|--------------|--------------|----------------|----------------|--------------|
| <i>Median MAFE's based on statistical releases</i>       |              |              |              |                |                |              |
| GDP and components                                       | 1.094        | 1.061        | <b>0.990</b> | 1.022          | <b>0.990</b>   | 0.994        |
| Housing  | 1.043        | 1.071        | 1.061        | 1.062          | 1.062          | <b>0.979</b> |
| IP   | 1.030        | 1.040        | <b>0.983</b> | 1.022          | 0.987          | 0.997        |
| Employment situation                                     | 1.188        | 1.029        | 1.004        | 1.075          | 1.004          | <b>0.937</b> |
| Productivity/Costs                                       | 1.320        | 1.544        | 1.112        | 1.397          | 1.109          | <b>0.891</b> |
| Assets/Liabilities of banks                              | <b>0.968</b> | 1.128        | 0.979        | 0.998          | 0.970          | 0.969        |
| Interest rates   | 1.624        | 1.050        | 1.002        | 1.113          | 1.000          | <b>0.969</b> |
| Money stock  | <b>0.933</b> | <b>0.933</b> | 0.935        | 0.934          | 0.938          | 1.019        |
| Currecny in circulation                                  | 0.916        | 0.920        | 0.918        | 0.925          | 0.920          | <b>0.884</b> |
| MZM  | 0.928        | 0.931        | 0.930        | <b>0.922</b>   | 0.934          | 1.117        |
| Money velocity   | 1.000        | 0.987        | <b>0.985</b> | 0.993          | 0.987          | 1.013        |
| Consumer Credit  | 0.967        | 1.006        | <b>0.964</b> | 0.990          | 0.968          | 1.015        |
| CPI  | 1.190        | 1.698        | <b>0.964</b> | 1.051          | 0.968          | 1.031        |
| PPI  | 0.983        | 0.964        | 0.964        | 0.989          | <b>0.962</b>   | 1.010        |
| Stock prices   | 0.968        | <b>0.948</b> | 0.952        | 0.951          | 0.959          | 1.031        |
| Exchange rates   | 1.093        | 1.053        | 1.029        | <b>1.027</b>   | 1.033          | 1.036        |
| ISM surveys  | 1.045        | 1.289        | <b>0.982</b> | 1.026          | 0.984          | 0.990        |
| <i>MAFE descriptives based on total number of series</i> |              |              |              |                |                |              |
| median   | 1.046        | 1.029        | <b>0.989</b> | 1.025          | <b>0.989</b>   | <b>0.989</b> |
| variance   | 0.693        | 1.776        | 0.009        | 0.081          | 0.007          | 0.009        |
| min  | 0.578        | 0.575        | 0.576        | 0.574          | 0.578          | 0.610        |
| max  | 6.707        | 9.883        | 1.609        | 3.386          | 1.364          | 1.275        |

Note: Entries are MAFE-based statistics relative to the MAFE of an AR(2) model.

Table 4: MSFE results for the five Bayes shrinkage estimators and the DFM,  $h = 1$

|  | Jef   | St-t         | LASSO        | Fused<br>LASSO | Elastic<br>Net | DFM          |
|--|-------|--------------|--------------|----------------|----------------|--------------|
| <i>Median MSFE's based on statistical releases</i>       |       |              |              |                |                |              |
| GDP and components                                       | 1.036 | 0.996        | 0.993        | 1.002          | <b>0.994</b>   | 1.047        |
| Housing  | 1.236 | 1.068        | 1.059        | 1.059          | 1.059          | <b>0.974</b> |
| IP   | 1.017 | 1.022        | 0.998        | 1.024          | 1.000          | <b>0.843</b> |
| Employment situation                                     | 1.036 | 1.026        | 1.019        | 1.037          | 1.019          | <b>0.899</b> |
| Productivity-Costs                                       | 1.003 | 1.108        | 0.986        | 1.023          | 0.985          | <b>0.896</b> |
| Assets-Liabilities of banks                              | 0.970 | 0.973        | 0.969        | 0.973          | 0.972          | <b>0.960</b> |
| Interest rates   | 1.045 | <b>0.954</b> | 0.959        | 0.995          | 0.956          | 0.995        |
| Money stock  | 0.948 | 0.949        | 0.945        | 0.946          | <b>0.943</b>   | 1.094        |
| Currecny in circulation                                  | 0.746 | 0.747        | 0.750        | 0.748          | 0.747          | <b>0.563</b> |
| MZM  | 0.920 | 0.905        | 0.907        | <b>0.904</b>   | <b>0.904</b>   | 1.668        |
| Money velocity   | 1.009 | 0.994        | 0.996        | 1.007          | <b>0.994</b>   | 1.129        |
| Consumer Credit  | 0.989 | 0.978        | <b>0.976</b> | 0.985          | <b>0.976</b>   | 1.087        |
| CPI  | 0.995 | 0.940        | <b>0.936</b> | 0.945          | <b>0.936</b>   | 0.945        |
| PPI  | 0.939 | 0.947        | 0.942        | 0.945          | 0.945          | <b>0.906</b> |
| Stock prices   | 1.007 | 1.007        | 1.006        | 1.006          | <b>1.003</b>   | 1.133        |
| Exchange rates   | 0.999 | 1.002        | 0.999        | 1.000          | <b>0.998</b>   | 1.016        |
| ISM surveys  | 1.091 | 1.236        | 1.026        | 1.016          | 1.025          | <b>0.932</b> |
| <i>MSFE descriptives based on total number of series</i> |       |              |              |                |                |              |
| median   | 1.007 | 0.994        | 0.991        | 1.002          | 0.991          | <b>0.958</b> |
| variance   | 0.041 | 0.084        | 0.006        | 0.010          | 0.006          | 0.054        |
| min  | 0.671 | 0.747        | 0.666        | 0.706          | 0.664          | 0.452        |
| max  | 2.380 | 2.797        | 1.474        | 1.545          | 1.370          | 2.498        |

Note: Entries are MSFE-based statistics relative to the MSFE of an AR(2) model.

Table 5: MSFE results for the five Bayes shrinkage estimators and the DFM,  $h = 2$

|  | Jef          | St-t  | LASSO        | Fused<br>LASSO | Elastic<br>Net | DFM          |
|--|--------------|-------|--------------|----------------|----------------|--------------|
| <i>Median MSFE's based on statistical releases</i>       |              |       |              |                |                |              |
| GDP and components                                       | 1.074        | 1.028 | 1.004        | 1.022          | <b>0.998</b>   | 1.004        |
| Housing  | 1.168        | 1.082 | 1.072        | 1.074          | <b>1.071</b>   | 1.161        |
| IP   | 1.041        | 1.152 | 1.012        | 1.049          | <b>1.011</b>   | 1.013        |
| Employment situation                                     | 1.055        | 1.035 | 1.020        | 1.036          | 1.022          | <b>0.985</b> |
| Productivity-Costs                                       | 1.128        | 1.327 | 1.075        | 1.155          | 1.077          | <b>0.882</b> |
| Assets-Liabilities of banks                              | 0.976        | 1.037 | 0.980        | 0.970          | 0.977          | <b>0.912</b> |
| Interest rates   | 1.260        | 1.008 | 0.994        | 1.084          | 0.998          | <b>0.946</b> |
| Money stock  | 0.974        | 0.974 | 0.976        | <b>0.972</b>   | 0.973          | 1.062        |
| Currecny in circulation                                  | 0.785        | 0.783 | 0.786        | 0.788          | 0.785          | <b>0.627</b> |
| MZM  | 0.915        | 0.911 | 0.907        | <b>0.904</b>   | 0.910          | 1.392        |
| Money velocity   | 1.021        | 1.003 | <b>1.002</b> | 1.019          | 1.005          | 1.026        |
| Consumer Credit  | 0.986        | 0.994 | <b>0.991</b> | 0.996          | 0.993          | 0.993        |
| CPI  | 1.048        | 0.967 | 0.966        | 0.986          | <b>0.965</b>   | 0.999        |
| PPI  | 0.979        | 0.968 | <b>0.967</b> | 0.976          | 0.968          | 0.975        |
| Stock prices   | <b>0.994</b> | 1.000 | 1.000        | 1.000          | 1.000          | 1.073        |
| Exchange rates   | 1.020        | 1.012 | 1.011        | <b>1.010</b>   | 1.011          | 1.023        |
| ISM surveys  | 1.036        | 1.697 | <b>1.007</b> | 1.020          | 1.010          | 1.043        |
| <i>MSFE descriptives based on total number of series</i> |              |       |              |                |                |              |
| median   | 1.035        | 1.018 | 0.998        | 1.017          | 0.999          | <b>0.995</b> |
| variance   | 0.144        | 0.205 | 0.007        | 0.017          | 0.006          | 0.026        |
| min  | 0.563        | 0.568 | 0.557        | 0.562          | 0.554          | 0.304        |
| max  | 3.392        | 3.536 | 1.420        | 1.667          | 1.313          | 1.897        |

Note: Entries are MSFE-based statistics relative to the MSFE of an AR(2) model.

Table 6: MSFE results for the five Bayes shrinkage estimators and the DFM,  $h = 4$

|  | Jef          | St-t         | LASSO        | Fused<br>LASSO | Elastic<br>Net | DFM          |
|--|--------------|--------------|--------------|----------------|----------------|--------------|
| <i>Median MSFE's based on statistical releases</i>       |              |              |              |                |                |              |
| GDP and components                                       | 1.061        | 1.057        | <b>0.997</b> | 1.022          | 0.999          | 1.004        |
| Housing  | <b>1.065</b> | 1.097        | 1.087        | 1.092          | 1.087          | 1.106        |
| IP   | 1.016        | 1.025        | 0.996        | 1.023          | 0.998          | <b>0.986</b> |
| Employment situation                                     | 1.112        | 1.025        | 1.006        | 1.048          | 1.005          | <b>0.916</b> |
| Productivity-Costs                                       | 1.222        | 1.614        | 1.099        | 1.289          | 1.093          | <b>0.815</b> |
| Assets-Liabilities of banks                              | 0.975        | 1.094        | 0.977        | 0.988          | 0.975          | <b>0.924</b> |
| Interest rates   | 1.571        | 1.084        | 1.001        | 1.096          | 1.003          | <b>0.957</b> |
| Money stock  | <b>0.973</b> | 0.974        | 0.975        | <b>0.973</b>   | 0.975          | 1.114        |
| Currecny in circulation                                  | 0.905        | 0.906        | 0.905        | 0.908          | 0.907          | <b>0.782</b> |
| MZM  | 0.932        | 0.934        | 0.934        | <b>0.927</b>   | 0.938          | 1.194        |
| Money velocity   | 1.020        | 0.999        | 0.996        | 1.013          | 0.998          | <b>0.987</b> |
| Consumer Credit  | <b>0.969</b> | 1.026        | 0.971        | 0.987          | 0.973          | 1.032        |
| CPI  | 1.085        | 1.354        | 0.969        | 1.014          | <b>0.968</b>   | 1.081        |
| PPI  | 0.972        | <b>0.962</b> | 0.964        | 0.977          | 0.963          | 0.999        |
| Stock prices   | 0.984        | 0.976        | 0.978        | 0.979          | 0.981          | 1.090        |
| Exchange rates   | 1.098        | 1.060        | <b>1.035</b> | <b>1.035</b>   | 1.036          | 1.062        |
| ISM surveys  | 1.052        | 1.521        | 0.998        | 1.032          | <b>0.996</b>   | 1.026        |
| <i>MSFE descriptives based on total number of series</i> |              |              |              |                |                |              |
| median   | 1.039        | 1.022        | 0.995        | 1.026          | 0.996          | <b>0.987</b> |
| variance   | 0.386        | 0.417        | 0.008        | 0.047          | 0.006          | 0.028        |
| min  | 0.620        | 0.610        | 0.611        | 0.612          | 0.612          | 0.414        |
| max  | 5.389        | 6.784        | 1.573        | 2.771          | 1.378          | 1.620        |

Note: Entries are MSFE-based statistics relative to the MSFE of an AR(2) model.



Table 7: Hit-rates of the five Bayes estimators, total no. of series

|                                      | Jeffreys' | Student-t | LASSO       | Fused LASSO | Elastic Net |
|--------------------------------------|-----------|-----------|-------------|-------------|-------------|
| <i>Hit rates, <math>h = 1</math></i> |           |           |             |             |             |
| % of lowest MAFE                     | 14.0      | 17.1      | <b>35.7</b> | 10.9        | 22.5        |
| % of lowest MSFE                     | 15.5      | 20.9      | <b>34.9</b> | 7.8         | 20.9        |
| % of highest APL                     | 0.8       | 2.3       | <b>62.0</b> | 7.8         | 27.1        |
| <i>Hit rates, <math>h = 2</math></i> |           |           |             |             |             |
| % of lowest MAFE                     | 18.6      | 13.2      | <b>34.1</b> | 11.6        | 22.5        |
| % of lowest MSFE                     | 17.1      | 20.9      | <b>28.7</b> | 11.6        | 21.7        |
| % of highest APL                     | 1.6       | 0.8       | <b>60.5</b> | 13.9        | 23.3        |
| <i>Hit rates, <math>h = 4</math></i> |           |           |             |             |             |
| % of lowest MAFE                     | 17.8      | 12.4      | <b>31.8</b> | 10.9        | 27.1        |
| % of lowest MSFE                     | 13.2      | 17.1      | <b>34.9</b> | 10.1        | 24.8        |
| % of highest APL                     | 0.0       | 0.0       | <b>55.8</b> | 8.5         | 35.7        |

Note: This table shows the proportion of times (over the 129 series being forecasted) that each estimator achieved the lowest value of the MAFE and MSFE statistics, and the highest value of the Average Predictive Likelihood (APL).

Table 8: Average similarity of Bayes forecasts,  $h = 1$ : correlation (lower left) and mean absolute difference of forecasts (upper right)

|             | Jeffreys' | Student-t | LASSO | Fused LASSO | Elastic Net |
|-------------|-----------|-----------|-------|-------------|-------------|
| Jef         |           | 0.088     | 0.037 | 0.032       | 0.037       |
| St-t        | 0.944     |           | 0.080 | 0.083       | 0.080       |
| LASSO       | 1.000     | 0.944     |       | 0.015       | 0.002       |
| Fused LASSO | 1.000     | 0.945     | 1.000 |             | 0.015       |
| Elastic Net | 1.000     | 0.944     | 1.000 | 1.000       |             |

Table 9: MAFE, MSFE and Predictive Likelihoods for all 129 series, orthogonal predictors,  $h = 1$

|  | Jeffreys' | Student-t | LASSO  | Fused LASSO | Elastic Net |
|--|-----------|-----------|--------|-------------|-------------|
| <i>MAFE descriptives based on total number of series</i> |           |           |        |             |             |
| median   | 0.9876    | 0.9871    | 0.9873 | 0.9846      | 0.9863      |
| 25% quantile   | 0.9539    | 0.9536    | 0.9523 | 0.9551      | 0.9505      |
| 75% quantile   | 1.0166    | 1.0128    | 1.0183 | 1.0139      | 1.0167      |
| variance   | 0.0061    | 0.0068    | 0.0060 | 0.0062      | 0.0061      |
| min  | 0.6768    | 0.6812    | 0.6813 | 0.6800      | 0.6821      |
| max  | 1.2988    | 1.3151    | 1.2797 | 1.3053      | 1.2803      |
| <i>MSFE descriptives based on total number of series</i> |           |           |        |             |             |
| median   | 0.9901    | 0.9923    | 0.9901 | 0.9903      | 0.9904      |
| 25% quantile   | 0.9558    | 0.9558    | 0.9559 | 0.9548      | 0.9510      |
| 75% quantile   | 1.0156    | 1.0227    | 1.0182 | 1.0178      | 1.0175      |
| variance   | 0.0046    | 0.0049    | 0.0046 | 0.0048      | 0.0047      |
| min  | 0.6714    | 0.6726    | 0.6666 | 0.6676      | 0.6670      |
| max  | 1.2197    | 1.2750    | 1.2150 | 1.2397      | 1.2160      |
| <i>PL descriptives based on total number of series</i>   |           |           |        |             |             |
| median   | 0.4190    | 0.4175    | 0.4711 | 0.4837      | 0.4724      |
| 25% quantile   | 0.2241    | 0.2232    | 0.2636 | 0.2717      | 0.2636      |
| 75% quantile   | 0.7112    | 0.7050    | 0.7839 | 0.7968      | 0.7854      |
| variance   | 0.2065    | 0.1679    | 0.1867 | 0.1954      | 0.1876      |
| min  | 0.0380    | 0.0379    | 0.0487 | 0.0510      | 0.0487      |
| max  | 3.7777    | 3.2578    | 3.3826 | 3.4491      | 3.3998      |

Table 10: LASSO forecasts of US GDP corr. predictors,  $h = 4$

|                         | LASSO 1 | LASSO 2 | LASSO 3 | LASSO 4 | DFM  |
|-------------------------|---------|---------|---------|---------|------|
| MAFE                    | 0.38    | 0.97    | 0.88    | 0.87    | 0.37 |
| MSFE                    | 0.24    | 0.93    | 0.81    | 0.79    | 0.22 |
| Corr.with DFM forecasts | 0.90    | 0.27    | 0.49    | 0.48    | 1    |

Note: The LASSO 1,2,3,4 models are the four univariate regressions described in the text, with estimation of  $\lambda$  using 1)  $(r, \delta) = (0.01, 0.01)$ , 2)  $(r, \delta) = (1, 0.1)$ , 3)  $(r, \delta) = (3, 1)$  and 4) marginal maximum likelihood.

## Recent titles

### CORE Discussion Papers

- 2010/68. Mathieu VAN VYVE. Fixed-charge transportation on a path: optimization, LP formulations and separation.
- 2010/69. Roland Iwan LUTTENS. Lower bounds rule!
- 2010/70. Fred SCHROYEN and Adekola OYENUGA. Optimal pricing and capacity choice for a public service under risk of interruption.
- 2010/71. Carlotta BALESTRA, Thierry BRECHET and Stéphane LAMBRECHT. Property rights with biological spillovers: when Hardin meets Meade.
- 2010/72. Olivier GERGAUD and Victor GINSBURGH. Success: talent, intelligence or beauty?
- 2010/73. Jean GABSZEWICZ, Victor GINSBURGH, Didier LAUSSEL and Shlomo WEBER. Foreign languages' acquisition: self learning and linguistic schools.
- 2010/74. Cédric CEULEMANS, Victor GINSBURGH and Patrick LEGROS. Rock and roll bands, (in)complete contracts and creativity.
- 2010/75. Nicolas GILLIS and François GLINEUR. Low-rank matrix approximation with weights or missing data is NP-hard.
- 2010/76. Ana MAULEON, Vincent VANNETELBOSCH and Cecilia VERGARI. Unions' relative concerns and strikes in wage bargaining.
- 2010/77. Ana MAULEON, Vincent VANNETELBOSCH and Cecilia VERGARI. Bargaining and delay in patent licensing.
- 2010/78. Jean J. GABSZEWICZ and Ornella TAROLA. Product innovation and market acquisition of firms.
- 2010/79. Michel LE BRETON, Juan D. MORENO-TERNERO, Alexei SAVVATEEV and Shlomo WEBER. Stability and fairness in models with a multiple membership.
- 2010/80. Juan D. MORENO-TERNERO. Voting over piece-wise linear tax methods.
- 2010/81. Jean HINDRIKS, Marijn VERSCHELDE, Glenn RAYP and Koen SCHOORS. School tracking, social segregation and educational opportunity: evidence from Belgium.
- 2010/82. Jean HINDRIKS, Marijn VERSCHELDE, Glenn RAYP and Koen SCHOORS. School autonomy and educational performance: within-country evidence.
- 2010/83. Dunia LOPEZ-PINTADO. Influence networks.
- 2010/84. Per AGRELL and Axel GAUTIER. A theory of soft capture.
- 2010/85. Per AGRELL and Roman KASPERZEC. Dynamic joint investments in supply chains under information asymmetry.
- 2010/86. Thierry BRECHET and Pierre M. PICARD. The economics of airport noise: how to manage markets for noise licenses.
- 2010/87. Eve RAMAEKERS. Fair allocation of indivisible goods among two agents.
- 2011/1. Yu. NESTEROV. Random gradient-free minimization of convex functions.
- 2011/2. Olivier DEVOLDER, François GLINEUR and Yu. NESTEROV. First-order methods of smooth convex optimization with inexact oracle.
- 2011/3. Luc BAUWENS, Gary KOOP, Dimitris KOROBILIS and Jeroen V.K. ROMBOUTS. A comparison of forecasting procedures for macroeconomic series: the contribution of structural break models.
- 2011/4. Taoufik BOUEZMARNI and Sébastien VAN BELLEGEM. Nonparametric Beta kernel estimator for long memory time series.
- 2011/5. Filippo L. CALCIANO. The complementarity foundations of industrial organization.
- 2011/6. Vincent BODART, Bertrand CANDELON and Jean-François CARPANTIER. Real exchanges rates in commodity producing countries: a reappraisal.
- 2011/7. Georg KIRCHSTEIGER, Marco MANTOVANI, Ana MAULEON and Vincent VANNETELBOSCH. Myopic or farsighted? An experiment on network formation.
- 2011/8. Florian MAYNERIS and Sandra PONCET. Export performance of Chinese domestic firms: the role of foreign export spillovers.
- 2011/9. Hiroshi UNO. Nested potentials and robust equilibria.

## Recent titles

### CORE Discussion Papers - continued

- 2011/10. Evgeny ZHELOBODKO, Sergey KOKOVIN, Mathieu PARENTI and Jacques-François THISSE. Monopolistic competition in general equilibrium: beyond the CES.
- 2011/11. Luc BAUWENS, Christian HAFNER and Diane PIERRET. Multivariate volatility modeling of electricity futures.
- 2011/12. Jacques-François THISSE. Geographical economics: a historical perspective.
- 2011/13. Luc BAUWENS, Arnaud DUFAYS and Jeroen V.K. ROMBOUTS. Marginal likelihood for Markov-switching and change-point GARCH models.
- 2011/14. Gilles GRANDJEAN. Risk-sharing networks and farsighted stability.
- 2011/15. Pedro CANTOS-SANCHEZ, Rafael MONER-COLONQUES, José J. SEMPERE-MONERRIS and Oscar ALVAREZ-SANJAIME. Vertical integration and exclusivities in maritime freight transport.
- 2011/16. Géraldine STRACK, Bernard FORTZ, Fouad RIANE and Mathieu VAN VYVE. Comparison of heuristic procedures for an integrated model for production and distribution planning in an environment of shared resources.
- 2011/17. Juan A. MAÑEZ, Rafael MONER-COLONQUES, José J. SEMPERE-MONERRIS and Amparo URBANO Price differentials among brands in retail distribution: product quality and service quality.
- 2011/18. Pierre M. PICARD and Bruno VAN POTTELSBERGHE DE LA POTTERIE. Patent office governance and patent system quality.
- 2011/19. Emmanuelle AURIOL and Pierre M. PICARD. A theory of BOT concession contracts.
- 2011/20. Fred SCHROYEN. Attitudes towards income risk in the presence of quantity constraints.
- 2011/21. Dimitris KOROBILIS. Hierarchical shrinkage priors for dynamic regressions with many predictors.

## Books

- P. VAN HENTENRYCKE and L. WOLSEY (eds.) (2007), *Integration of AI and OR techniques in constraint programming for combinatorial optimization problems*. Berlin, Springer.
- P-P. COMBES, Th. MAYER and J-F. THISSE (eds.) (2008), *Economic geography: the integration of regions and nations*. Princeton, Princeton University Press.
- J. HINDRIKS (ed.) (2008), *Au-delà de Copernic: de la confusion au consensus ?* Brussels, Academic and Scientific Publishers.
- J-M. HURIOT and J-F. THISSE (eds) (2009), *Economics of cities*. Cambridge, Cambridge University Press.
- P. BELLEFLAMME and M. PEITZ (eds) (2010), *Industrial organization: markets and strategies*. Cambridge University Press.
- M. JUNGER, Th. LIEBLING, D. NADDEF, G. NEMHAUSER, W. PULLEYBLANK, G. REINELT, G. RINALDI and L. WOLSEY (eds) (2010), *50 years of integer programming, 1958-2008: from the early years to the state-of-the-art*. Berlin Springer.
- G. DURANTON, Ph. MARTIN, Th. MAYER and F. MAYNERIS (eds) (2010), *The economics of clusters – Lessons from the French experience*. Oxford University Press.
- J. HINDRIKS and I. VAN DE CLOOT (eds) (2011), *Notre pension en héritage*. Itinera Institute.

## CORE Lecture Series

- D. BIENSTOCK (2001), Potential function methods for approximately solving linear programming problems: theory and practice.
- R. AMIR (2002), Supermodularity and complementarity in economics.
- R. WEISMANTEL (2006), Lectures on mixed nonlinear programming.